



TUGAS AKHIR - SS141501

**PENGUNAAN METODE *CLASSIFICATION AND REGRESSION TREES* (CART) UNTUK KLASIFIKASI
REKURENSI PASIEN KANKER SERVIKS
DI RSUD DR. SOETOMO SURABAYA**

Siti Holis Sumartini
NRP 1313 105 013

Dosen Pembimbing
Santi Wulan Purnami, M.Si, Ph.D

Program Studi S1 Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT- SS141501

***CLASSIFICATION AND REGRESSION TREES
(CART) METHOD FOR CLASSIFICATION OF
CERVICAL CANCER PATIENTS RECURRENCE
AT DR. SOETOMO HOSPITAL***

Siti Holis Sumartini
NRP 1313 105 013

Supervisor
Santi Wulan Purnami, M.Si, Ph.D

Undergraduate Programme of Statistics
Faculty of Mathematics and Natural Science
Institut Teknologi Sepuluh Nopember
Surabaya 2015

LEMBAR PENGESAHAN

PENGUNAAN METODE *CLASSIFICATION AND REGRESSION TREES* (CART) UNTUK KLASIFIKASI REKURENSI PASIEN KANKER SERVIKS DI RSUD DR. SOETOMO SURABAYA

TUGAS AKHIR

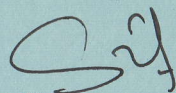
Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

SITI HOLIS SUMARTINI
NRP. 1313 105 013

Disetujui oleh Pembimbing Tugas Akhir :

Santi Wulan Purnami, M.Si, Ph.D.
NIP. 19720923 199803 2 001



Mengetahui
Ketua Jurusan Statistika FMIPA-ITS



Dr. Muhammad Mashuri, MT.
NIP. 19620408 198701 1 001

SURABAYA, Juli 2015

PENGUNAAN METODE *CLASSIFICATION AND REGRESSION TREES* (CART) UNTUK KLASIFIKASI REKURENSI PASIEN KANKER SERVIKS DI RSUD DR. SOETOMO SURABAYA

Nama Mahasiswa : Siti Holis Sumartini
NRP : 1313 105 013
Jurusan : Statistika
Pembimbing : Santi Wulan Purnami, M.Si, Ph.D.

Abstrak

Kanker serviks merupakan kanker yang menyerang area bawah rahim. Pengobatan kanker serviks tergantung pada besarnya ukuran dan stadium kanker. Kasus rekurensi sering terjadi walaupun pengobatan awal telah dilakukan. Salah satu rumah sakit yang menyediakan fasilitas terapi untuk pasien kanker serviks adalah RSUD dr. Soetomo Surabaya. Permasalahannya adalah bagaimana klasifikasi pasien kanker serviks di RSUD dr. Soetomo Surabaya yang rekuren dan tidak rekuren berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks dan mengetahui faktor-faktor yang berpengaruh terhadap rekurensi kanker serviks. Rekurensi yang dimaksudkan di penelitian ini adalah kembalinya pasien kanker serviks ke RSUD dr. Soetomo karena penyakit yang sama. Data yang digunakan merupakan data sekunder, yang diperoleh dari rekam medis pasien kanker serviks di RSUD dr. Soetomo Surabaya pada tahun 2014 dengan jumlah data sebanyak 810 pasien. Berdasarkan analisis yang telah dilakukan, pasien yang rekuren lebih banyak dibandingkan pasien yang tidak rekuren dengan persentase sebesar 57,78 persen untuk yang rekuren. Klasifikasi CART menghasilkan bahwa variabel yang paling berpengaruh terhadap rekurensi kanker serviks adalah variabel jenis pengobatan yang dijalani oleh pasien, selain itu variabel usia, status anemia dan status penyakit penyerta juga berpengaruh terhadap rekurensi kanker serviks. Ketepatan klasifikasi yang diperoleh untuk data prediksi sebesar 69,14%.

Kata kunci : klasifikasi, CART, kanker serviks, rekurensi

Halaman ini sengaja dikosongkan

CLASSIFICATION AND REGRESSION TREES (CART) METHOD FOR CLASSIFICATION OF CERVICAL CANCER PATIENTS RECURRENCE AT DR. SOETOMO HOSPITAL

Name of Student : Siti Holis Sumartini
NRP : 1310 030 007
Department : Statistics
Supervisor : Santi Wulan Purnami, M.Si, Ph.D.

Abstract

Cervical cancer is a cancer that attacks the lower area of the uterus. Treatment of cervical cancer depends on the size and stage of the cancer. Cases of recurrence often occurs despite initial treatment has been carried out. One of the hospitals that provide treatment for patients with cervical cancer is dr. Soetomo hospital. The problem is how the classification of cervical cancer patients in dr. Soetomo recurrent and recurrent based on the factors that affect recurrence of cervical cancer and determine the factors that influence the recurrence of cervical cancer. Recurrence is intended in this study is the return of cervical cancer patients to dr. Soetomo of the same disease. The data used is secondary data obtained from the medical records of patients with cervical cancer at dr. Soetomo hospital in 2014 with the amount of data as much as 810 patients. Based on the analysis that has been done, patients who are more recurrences compared patients who no recurrences with percentage of 57,78% recurrences. Classification CART produce that the most influential variables on recurrence of cervical cancer is a variable type of treatment undergone by the patient, in addition to the variables of age, anemia status and comorbidities status also affects the cervical cancer recurrences. Classification accuracy obtained for data prediction is 69,14%.

Key Word : classification, CART, cervical cancer, recurrence

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Assalamualaikum Wr. Wb.

Segala puji dan syukur kehadiran Allah SWT. yang telah memberikan rahmat dan hidayah-Nya, serta shalawat dan salam senantiasa tercurahkan bagi Nabi Muhammad SAW. Beserta sahabat dan keluarganya, sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul

“PENGGUNAAN METODE *CLASSIFICATION AND REGRESSION TREES* (CART) UNTUK KLASIFIKASI REKURENSI PASIEN KANKER SERVIKS DI RSUD DR. SOETOMO SURABAYA”

Selesainya laporan Tugas Akhir serta laporan ini tidak lepas dari peranan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Muhammad Mashuri, M.T selaku Ketua Jurusan Statistika FMIPA ITS.
2. Ibu Dra. Lucia Aridinanti, M.T selaku Ketua Program Studi (Kaprodi) S-1 Jurusan Statistika FMIPA ITS.
3. Ibu Santi Wulan Purnami, M.Si, Ph.D selaku dosen pembimbing yang telah membimbing dan mengarahkan penulis selama penyusunan laporan Tugas Akhir.
4. Bapak Dr. Bambang Widjanarko Otok, S.Si, M.Si dan ibu Shofi Andari, M.Si selaku dosen penguji, terima kasih telah memberikan arahan dan masukan kepada penulis dalam menyelesaikan laporan Tugas Akhir ini.
5. Ibu Dr. Vita Ratnasari, S.Si, M.Si selaku dosen wali yang telah memberikan bimbingan dan semangat kepada penulis selama kuliah di Statistika ITS.
6. Ayahanda Abdul Halik (alm) dan ibunda Badiyah tercinta beserta saudara-saudara penulis, mas Faqih, mas Rachman dan mas Kusnadi, serta nenek penulis yang selalu memberikan

doa, semangat dan dukungan selama ini sehingga penulis dapat sekolah hingga bangku perkuliahan.

7. Sahabat-sahabat yang penulis sayangi, Ayuk, Pitri, Yaumil, Anin, Alfiana, Ella, Lia dan Linda yang selalu setia memberikan dukungan dan motivasi kepada penulis.
8. Mbak Nariza dan Kurnia yang menjadi partner setia ketika mengentry data.
9. Teman-teman seperjuangan TA demi Toga 112 atas kebersamaan dalam menyelesaikan Tugas Akhir, khususnya teman-teman Lintas Jalur S1 Jurusan Statistika 2013 yang saling memberikan motivasi dan dukungan dalam proses penyelesaian Tugas Akhir.
10. Pihak-pihak yang sudah banyak membantu penulis dalam proses pengerjaan Tugas Akhir ini, yang tidak dapat penulis sebutkan satu per satu.

Semoga kebaikan dan bantuan yang telah diberikan kepada penulis dibalas dengan kebaikan yang lebih oleh Allah SWT.

Penulis menyadari bahwa laporan Tugas Akhir ini masih terdapat kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat diharapkan. Semoga laporan Tugas Akhir ini dapat memberikan manfaat baik bagi penulis, pembaca dan semua pihak.

Wassalamualaikum Wr.Wb.

Surabaya, Juli 2015

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB I. PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Manfaat	5
1.5 Batasan Masalah	5
BAB II. TINJAUAN PUSTAKA	
2.1 <i>Classification And Regression Trees (CART)</i>	7
2.1.1 Pembentukan Pohon Klasifikasi	8
2.1.2 Pemangkasan Pohon Klasifikasi	12
2.1.3 Penentuan Pohon Klasifikasi Optimal	13
2.1.4 Ilustrasi CART	14
2.2 Ukuran Ketepatan Klasifikasi	23
2.3 <i>Cross Validation</i>	25
2.4 Gambaran Mengenai Kanker Serviks	26
2.4.1 Faktor Penyebab Kanker Serviks	27
2.4.2 Gejala dan Pengobatan Kanker Serviks	28
2.4.3 Stadium Kanker Serviks	29
2.4.4 Rekurensi Kanker Serviks.....	31
2.4.5 Faktor-faktor Penyebab Rekurensi Kanker Serviks	31
BAB III. METODOLOGI PENELITIAN	
3.1 Sumber Data.....	33

3.2 Variabel Penelitian	33
3.3 Metode Analisis Data	34
BAB IV. ANALISIS DAN PEMBAHASAN	
4.1 Deskripsi Karakteristik Pasien	37
4.1.1 Karakteristik Pasien Berdasarkan Usia Pasien ..	38
4.1.2 Karakteristik Pasien Berdasarkan Faktor Status Anemia	38
4.1.3 Karakteristik Pasien Berdasarkan Faktor Stadium Kanker Serviks	40
4.1.4 Karakteristik Pasien Berdasarkan Jenis Pengobatan yang Dijalani Pasien	41
4.1.5 Karakteristik Pasien Berdasarkan Status Komplikasi	43
4.1.6 Karakteristik Pasien Berdasarkan Status Penyakit Penyerta	45
4.2 Analisis CART untuk Klasifikasi Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya	46
4.2.1 Pembentukan Pohon Klasifikasi Maksimal	47
4.2.2 Pemangkasan Pohon Klasifikasi Maksimal (<i>Prunning</i>).....	49
4.2.3 Pemilihan Pohon Klasifikasi Optimal	50
4.2.4 Hasil Ketepatan Klasifikasi Pohon CART	54
BAB V. KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	59
5.2 Saran.....	60
DAFTAR PUSTAKA	63
LAMPIRAN	67

DAFTAR TABEL

	Halaman
Tabel 2.1 Variabel Ilustrasi CART	14
Tabel 2.2 Data Sampel untuk Ilustrasi CART	14
Tabel 2.3 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar 40	16
Tabel 2.4 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar 51	16
Tabel 2.5 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar 58,5	16
Tabel 2.6 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> sebesar 69,5	16
Tabel 2.7 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 1:2&3	17
Tabel 2.8 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 2:1&3	17
Tabel 2.9 Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 3:1&2	18
Tabel 2.10 Perbandingan <i>Goodness of Split</i> dari Setiap Kemungkinan Pemilah 1	18
Tabel 2.11 Perbandingan <i>Goodness of Split</i> dari Setiap Kemungkinan Pemilah 2	19
Tabel 2.12 Perbandingan <i>Goodness of Split</i> dari Setiap Kemungkinan Pemilah 3	21
Tabel 2.13 Perhitungan Peluang Klasifikasi Simpul	22
Tabel 2.14 <i>Crosstab</i> Ketepatan Klasifikasi	24
Tabel 2.15 Data Eksperimen Untuk 5-Fold Cross Validation	26
Tabel 2.16 Stadium Kanker Serviks	30
Tabel 3.1 Struktur Data	33
Tabel 3.2 Variabel Penelitian.....	34
Tabel 4.1 Deskripsi Usia Pasien Berdasarkan Status Rekurensi	38
Tabel 4.2 Deskripsi Status Anemia Berdasarkan Status Rekurensi	39

Tabel 4.3	Deskripsi Stadium Berdasarkan Status Rekurensi	41
Tabel 4.4	Deskripsi Jenis Pengobatan yang Dijalani Pasien Berdasarkan Status Rekurensi	42
Tabel 4.5	Deskripsi Status Komplikasi Berdasarkan Status Rekurensi	44
Tabel 4.6	Deskripsi Status Penyakit Penyerta Berdasarkan Status Rekurensi	45
Tabel 4.7	Perhitungan Pemilah pada Setiap Variabel	47
Tabel 4.8	Skor Variabel Terpenting dari Pohon Klasifikasi Maksimal	48
Tabel 4.9	Urutan Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>)	50
Tabel 4.10	Skor Variabel Terpenting dari Pohon Klasifikasi Optimal	52
Tabel 4.11	Karakteristik Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya	54
Tabel 4.12	Klasifikasi Data <i>Learning</i> Menggunakan <i>5-Fold CV</i>	55
Tabel 4.13	Evaluasi Performansi pada Data <i>Learning</i>	56
Tabel 4.14	Ketepatan Klasifikasi untuk Data Prediksi	56

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Struktur Pohon Klasifikasi	7
Gambar 2.2 Pohon Klasifikasi Hasil Pemilahan Simpul Awal	19
Gambar 2.3 Pohon Klasifikasi Hasil Pemilahan Simpul 2	20
Gambar 2.4 Pohon Klasifikasi Hasil Pemilahan Simpul 3	21
Gambar 2.5 Pohon Klasifikasi Setelah Dilakukan Pelabelan Kelas	23
Gambar 3.1 Diagram Alir Analisis Data	36
Gambar 4.1 Deskripsi Status Rekurensi Kanker Serviks	37
Gambar 4.2 Deskripsi Status Anemia	39
Gambar 4.3 Deskripsi Stadium Kanker Serviks	40
Gambar 4.4 Deskripsi Jenis Pengobatan yang Dijalani Pasien	42
Gambar 4.5 Deskripsi Status Komplikasi	43
Gambar 4.6 Deskripsi Status Penyakit Penyerta	45
Gambar 4.7 Konstruksi Pohon Klasifikasi Maksimal	48
Gambar 4.8 Plot <i>Relative Cost</i> dan Banyaknya <i>Terminal Nodes</i>	49
Gambar 4.9 Pohon Klasifikasi Optimal	51

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

	Halaman
Lampiran A Data Kanker Serviks	67
Lampiran B <i>Output</i> CART Hasil Klasifikasi	68
Lampiran C <i>Output Terminal Nodes</i> Untuk Klasifikasi Data <i>Learning</i>	74

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan salah satu penyebab utama kematian dibanyak negara terutama negara berkembang (WHO, 2010). Kanker merupakan suatu penyakit yang ditimbulkan oleh sel tunggal yang tumbuh tidak normal dan tidak terkendali sehingga dapat menjadi tumor ganas yang dapat menghancurkan dan merusak sel atau jaringan sehat. Pertumbuhan sel yang abnormal ini dapat menyebar ke berbagai organ dalam tubuh salah satunya pada area bawah rahim, yaitu daerah pada organ reproduksi wanita yang merupakan pintu masuk ke area rahim. Letaknya antara rahim (uterus) dan vagina yang menghubungkan rahim sebelah atas dengan vagina atau lebih dikenal dengan sebutan kanker mulut rahim (*serviks*) (Nurlaila dan Hadi, 2014).

Penyebab utama kanker serviks (kanker leher rahim) adalah HPV (*human papilloma virus*) atau virus papiloma manusia. Virus ini ditemukan pada 77% perempuan yang terjangkit HIV positif dan sisanya terhadap wanita siapa saja (Almarogi, 2015). Bahkan kanker ini akan menyerang siapa saja dalam rentan usia yang tak terbatas. Sebagian besar perempuan penderita kanker serviks ini tidak menyadari ketika HPV menyeranginya. Dan biasanya akan menimbulkan keluhan atau gejala saat penyakit sudah berada pada stadium lanjut. Penyebab banyak kematian pada kaum wanita adalah virus HPV tipe 16 dan 18. Virus ini sangat mudah berpindah dan menyebar, tidak hanya melalui cairan, tapi juga bisa berpindah melalui sentuhan kulit. Perjalanan dari infeksi virus menjadi kanker membutuhkan waktu cukup lama, sekitar 10-20 tahun (Meilinda, 2014).

Pengobatan kanker serviks tergantung pada besarnya ukuran kanker dan stadium kanker. Pada stadium awal, pengobatan kanker serviks dilakukan dengan cara menyingkirkan bagian yang sudah terkena kanker. Misalnya dengan melakukan pembedahan listrik, laser atau *cryosurgery* (membekukan dan

membuang jaringan abnormal). Sedangkan untuk pengobatan kanker serviks stadium lanjut, pengobatan kanker serviks dilakukan dengan terapi kemoterapi dan radioterapi. Pada stadium akhir atau kasus yang sudah parah, maka terpaksa harus dilakukan histerektomi, yaitu pembedahan pengangkatan rahim (uterus) secara total agar sel-sel kanker yang telah berkembang dalam kandungan tidak menyebar ke bagian lain dalam tubuh (Chamim, 2006).

Namun, pengobatan pada stadium awal tidak menjamin dapat menyembuhkan kanker serviks. Kasus rekurensi sering terjadi setelah pengobatan awal dilakukan. Rekurensi kanker serviks merupakan masalah yang patut diwaspadai walaupun pembedahan dan kemoradioterapi (gabungan dari kemoterapi dan radioterapi) telah mampu mengobati 80-95% lesi (keadaan jaringan yang abnormal pada tubuh) kanker derajat rendah (stadium 1 dan 2) dan 60% stadium 3. Deteksi dini pada kasus kekambuhan kanker serviks ini dapat meminimalkan resiko kanker serviks. Rekurensi yang dimaksudkan dalam penelitian ini adalah kembalinya pasien ke RSUD dr. Soetomo Surabaya setelah melakukan perawatan sebelumnya. Terdapat berbagai macam faktor risiko yang diduga memicu rekurensi kanker serviks, antara lain usia pasien, status anemia, stadium kanker serviks serta terapi yang dijalani oleh pasien. Oleh karena itu, penelitian ini bertujuan untuk menganalisis pengklasifikasian pasien kanker serviks berdasarkan faktor-faktor yang berpengaruh terhadap rekurensi kanker serviks dan mengetahui faktor-faktor apakah yang berpengaruh terhadap rekurensi kanker serviks. Terdapat berbagai macam metode klasifikasi, yaitu CART (*Classification And Regression Trees*), analisis regresi logistik, CHAID (*Chi-Square Automatic Interaction Detection*), SVM dan metode klasifikasi lainnya. Dalam penelitian ini, metode yang digunakan adalah klasifikasi CART karena dapat mengklasifikasikan data baru secara efisien.

Metode CART (*classification and regression trees*) merupakan suatu metode statistika nonparametrik yang dapat

menggambarkan hubungan antara variabel respon (dependen) dengan satu atau lebih variabel prediktor (variabel independen). Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Metode CART mempunyai beberapa kelebihan, CART merupakan metode nonparametrik sehingga tidak ada asumsi distribusi variabel prediktor yang perlu dipenuhi. Kedua, CART tidak hanya memberikan prediksi kelas, namun juga estimasi probabilitas kesalahan pengklasifikasian. Ketiga, metode ini memudahkan dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multivariabel karena struktur data dapat dilihat secara visual. Keempat, hasil klasifikasi akhir berbentuk sederhana dan mengklasifikasikan data baru secara efisien. Kelima, kemudahan dalam menginterpretasi hasil (Breiman, Friedman, Olshen, & Stone, 1993).

RSUD (Rumah Sakit Umum Daerah) dr. Soetomo Surabaya merupakan rumah sakit yang menjadi pusat rujukan terutama untuk kawasan Indonesia Timur, karena jumlah pasien yang berobat di rumah sakit milik Pemerintah Propinsi Jawa timur ini semakin bertambah banyak jumlahnya. Selain itu, rumah sakit ini merupakan salah satu rumah sakit yang menyediakan layanan kesehatan yang menyediakan fasilitas terapi untuk pasien kanker serviks (Mauliddah, 2014). Menurut Dewi (2013), Poli Onkologi Satu Atap (POSA) RSUD dr. Soetomo Surabaya tahun 2013 menunjukkan bahwa kanker serviks menempati urutan kedua terbanyak di RSUD dr. Soetomo Surabaya dan mempunyai kecenderungan meningkat dari tahun ke tahun.

Penelitian dengan menggunakan klasifikasi CART pernah dilakukan oleh Nuriyah (2013) yaitu membandingkan metode CHAID dan CART dalam menentukan klasifikasi alumni UIN Sunan Kalijaga berdasarkan masa Study dan dapat disimpulkan bahwa metode CART lebih efisien dalam mengklasifikasikan alumni UIN Sunan Kalijaga, hal ini dapat dilihat dari ketepatan klasifikasi CHAID sebesar 63.3 persen dan ketepatan klasifikasi CART sebesar 66.6 persen. Selain itu, klasifikasi CART pernah

dilakukan oleh Karisma (2013) untuk melakukan analisis klasifikasi pada penderita *diabetes mellitus* dengan Metode CHAID dan CART dan dapat disimpulkan bahwa berdasarkan metode CHAID diperoleh ketepatan klasifikasi sebesar 96.8 persen dengan variabel yang signifikan adalah obesitas, jenis kelamin, dan riwayat keluarga, sedangkan metode CART menghasilkan ketepatan klasifikasi sebesar 94.8 persen dan variabel yang berpengaruh adalah riwayat keluarga, obesitas, usia, dan jenis kelamin. Sedangkan penelitian mengenai rekurensi kanker serviks dilakukan oleh Novitasari (2014) mengenai Analisis Survival pada Data Rekurensi dengan Menggunakan *Counting Process Approach* dan Model PWP-GT dan diperoleh kesimpulan bahwa variabel yang berpengaruh terhadap rekurensi kanker serviks adalah jenis kanker dan stadium kanker serviks.

1.2 Rumusan Masalah

Kasus rekurensi kanker serviks sering terjadi setelah pengobatan kanker serviks dilakukan. Terdapat beberapa faktor risiko yang diduga berpengaruh terhadap rekurensi kanker serviks, yaitu usia, status anemia, stadium, jenis pengobatan, status komplikasi dan status penyakit penyerta. Salah satu rumah sakit yang menyediakan fasilitas terapi untuk pasien kanker serviks adalah RSUD dr. Soetomo Surabaya. Permasalahannya adalah bagaimana klasifikasi pasien kanker serviks di RSUD dr. Soetomo Surabaya yang rekuren dan tidak rekuren berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks dan mengetahui faktor-faktor yang berpengaruh terhadap rekurensi kanker serviks menggunakan metode CART

1.3 Tujuan

Adapun tujuan yang ingin dicapai pada penelitian ini adalah

1. Mengetahui faktor-faktor yang mempengaruhi rekurensi kanker serviks di RSUD dr. Soetomo Surabaya.

2. Menentukan klasifikasi pasien kanker serviks berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks di RSUD dr. Soetomo Surabaya.

1.4 Manfaat

Manfaat yang akan diperoleh dari penelitian ini adalah sebagai berikut.

1. Bagi RSUD dr. Soetomo Surabaya, penelitian ini diharapkan dapat digunakan untuk membantu dan mempermudah para dokter dalam menangani rekurensi kanker serviks, selain itu dapat juga mengetahui faktor-faktor yang mempengaruhi rekurensi kanker serviks.
2. Bagi institusi pendidikan, penelitian ini bagi diharapkan dapat digunakan untuk menambah wawasan dalam ilmu kesehatan, selain itu juga dapat dijadikan sebagai bahan pengetahuan untuk penelitian selanjutnya yang lebih mendalam mengenai kanker serviks dan juga kaitannya dengan metode CART.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah data yang digunakan merupakan data pasien kanker serviks di RSUD dr. Soetomo Surabaya yang melakukan rawat inap pada tahun 2014. Data yang diambil adalah data kedatangan kedua pasien ke RSUD dr. Soetomo Surabaya.

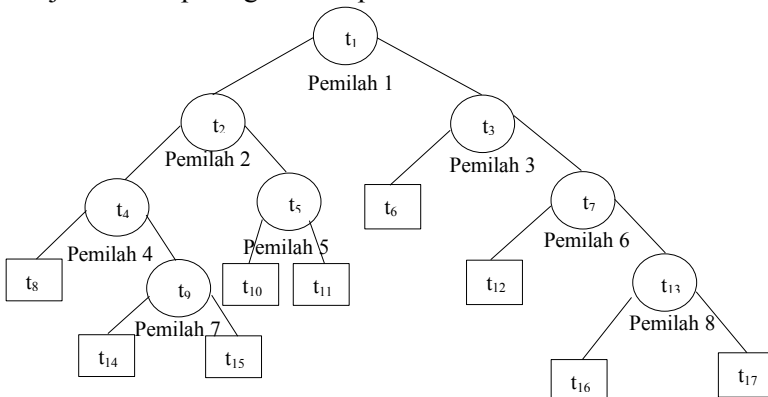
Halaman ini Sengaja Dikosongkan

BAB II

TINJAUAN PUSTAKA

2.1 *Classification and Regression Trees (CART)*

Classification and Regression Trees (CART) merupakan salah satu metode atau algoritma dari teknik pohon keputusan (*decision tree*). Metode yang dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone ini merupakan teknik klasifikasi dengan menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*) (Roger dan Lewis, 2000). Istilah “*binary*” berarti pemilahan dilakukan pada sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul (*node*) menjadi dua kelompok yang disebut simpul anak (*child nodes*). Istilah “*recursive*” berarti prosedur penyekatan secara biner dilakukan secara berulang-ulang. Setiap simpul anak yang diperoleh dari penyekatan simpul awal kemudian bisa dipilah kembali menjadi dua simpul anak lagi, dan begitu seterusnya hingga memenuhi kriteria tertentu. Sedangkan istilah “*partitioning*” memiliki arti bahwa proses klasifikasi dilakukan dengan cara memilah suatu kumpulan data menjadi beberapa bagian atau partisi.



Sumber: Breiman L., Friedman J.H., Olshen R.A., & Stone C.J. (1993). *Classification and Regression Trees (CART)*

Gambar 2.1 Struktur Pohon Klasifikasi

Ilustrasi dari struktur pohon klasifikasi ditunjukkan pada Gambar 2.1. Simpul awal yang merupakan variabel terpenting dalam kelas amatan disebut *parent node* dengan notasi simpul t_1 , simpul dalam (*internal nodes*) dinotasikan dengan t_2, t_3, t_4, t_7, t_9 dan t_{10} , serta simpul akhir (*terminal nodes*) yang dinotasikan dengan $t_5, t_6, t_8, t_{11}, t_{12}, t_{13}, t_{14}$ dan t_{15} dimana setelahnya tidak ada lagi pemilahan. Penghitungan *depth* (kedalaman) pohon dimulai dari simpul utama t_1 yang berada pada kedalaman 1, sedangkan t_2 dan t_3 berada pada kedalaman 2 begitu seterusnya hingga t_{12}, t_{13}, t_{14} dan t_{15} yang berada pada kedalaman 5. Selain itu, setiap simpul terminal diberi tanda dengan label kelas.

Menurut Breiman, et al (1993), CART akan menghasilkan pohon klasifikasi jika variabel respon mempunyai skala kategorik dan akan menghasilkan pohon regresi jika variabel respon berupa data kontinu. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Metode pengklasifikasian CART memiliki beberapa kelebihan. Pertama, CART merupakan metode nonparametrik sehingga tidak ada asumsi distribusi variabel prediktor yang perlu dipenuhi. Kedua, CART tidak hanya memberikan klasifikasi, namun juga estimasi probabilitas kesalahan pengklasifikasian. Ketiga, metode ini memudahkan dalam hal eksplorasi dan pengambilan keputusan pada struktur data yang kompleks dan multivariabel karena struktur data dapat dilihat secara visual. Keempat, hasil klasifikasi akhir berbentuk sederhana dan mengklasifikasikan data baru secara efisien. Kelima, kemudahan dalam menginterpretasi hasil.

Algoritma CART melalui tiga tahapan, yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi dan penentuan pohon klasifikasi optimum.

2.1.1 Pembentukan Pohon Klasifikasi

Pembentukan pohon klasifikasi diawali dengan menentukan variabel dan nilai dari variabel tersebut (*threshold*) untuk dijadikan pemilah tiap simpul. Dalam prosesnya, pembentukan pohon klasifikasi dibutuhkan data *learning* sampel

L yang terdiri atas N pengamatan. Menurut Breiman, et al (1993), proses pembentukan pohon klasifikasi terdiri atas 3 tahapan yaitu sebagai berikut.

a. Pemilihan Pemilah

Pada tahap ini, data yang digunakan adalah sampel data *learning* L yang kemudian dipilah berdasarkan aturan pemilahan dan kriteria *goodness of split*. Himpunan bagian yang dihasilkan dari proses pemilahan harus lebih homogen dibandingkan pemilahan sebelumnya. Hal ini dilakukan dengan mendefinisikan keheterogenan simpul (*impurity* atau $i(t)$). Menurut Breiman, et al (1993), fungsi keheterogenan yang sangat mudah dan sesuai diterapkan dalam berbagai kasus adalah Indeks Gini. Indeks Gini akan selalu memisahkan kelas dengan anggota paling besar atau kelas terpenting dalam simpul tersebut terlebih dahulu. Pemilahan yang memberikan nilai penurunan keheterogenan tertinggi merupakan pemilahan terbaik. Fungsi Indeks Gini dituliskan dalam persamaan berikut.

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j \quad (2.1)$$

Dengan $p(j|t)$ adalah proporsi kelas j pada simpul t dan $p(i|t)$ adalah proporsi kelas i pada simpul t .

Pemilahan simpul dimulai dengan memeriksa nilai-nilai variabel independen dan dilakukan secara rekursif pada setiap simpul dengan dua tahapan. Tahapan yang pertama adalah mencari semua kemungkinan pemilah pada variabel prediktor. Menurut Breiman, et al (1993), proses pemilahan simpul menjadi dua simpul anak dilakukan dengan mengikuti aturan sebagai berikut.

1. Setiap pemilahan hanya bergantung pada nilai yang berasal dari satu variabel prediktor saja.
2. Apabila variabel prediktor berskala kontinu, maka pemilahan yang diperbolehkan adalah $x_j \leq c_i$ dan $x_j > c_i$, dengan $i = 1, 2, \dots, n-1$ dan c_i adalah nilai tengah dari dua nilai amatan sampel berurutan yang berbeda dari variabel X_j . jika suatu ruang sampel berukuran n dan terdapat n nilai amatan sampel yang

berbeda pada variabel X_j , maka akan terdapat sebanyak $n-1$ kemungkinan pemilahan yang berbeda.

3. Apabila variabel prediktor berskala kategorik, maka pemilahan berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya dua simpul yang saling lepas (*disjoint*). Apabila variabel prediktor berskala nominal bertaraf L , maka akan diperoleh sebanyak $2^{L-1}-1$ pemilahan yang mungkin. Akan tetapi, apabila kategori variabel prediktor berskala ordinal bertaraf L , maka akan diperoleh sebanyak $L-1$ pemilahan yang mungkin.

Pemilahan yang terpilih akan membentuk suatu himpunan kelas yang disebut sebagai simpul. Simpul tersebut akan melakukan pemilahan secara rekursif sampai diperoleh simpul akhir (*terminal nodes*).

Setelah dilakukan pemilahan dari semua kemungkinan pemilah, maka tahapan berikutnya adalah menentukan kriteria *goodness of split* ($\phi(s,t)$) untuk mengevaluasi pemilah dari pemilah s pada simpul t . *Goodness of split* ($\phi(s,t)$) merupakan penurunan heterogenitas, yaitu.

$$\phi(s,t) = \Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2.2)$$

Dengan

$i(t)$: fungsi heterogenitas pada simpul t

p_L : proporsi pengamatan menuju simpul kiri

p_R : proporsi pengamatan menuju simpul kanan

$i(t_L)$: fungsi heterogenitas pada simpul anak kiri

$i(t_R)$: fungsi heterogenitas pada simpul anak kanan

Pemilah yang menghasilkan $\phi(s,t)$ lebih tinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi. Pengembangan pohon ini dilakukan dengan pencarian pemilah yang mungkin pada simpul t_1 yang kemudian akan dipilah menjadi t_2 dan t_3 oleh pemilah s , dan seterusnya. t_L dan t_R merupakan partisi dari simpul t menjadi dua himpunan bagian saling lepas dimana p_L dan p_R adalah proporsi masing-masing peluang simpul. Karena $t_L \cup t_R = t$ maka nilai $\Delta i(s,t)$ merepresentasikan perubahan dari kehetoregenan dalam simpul t

yang semata-mata disebabkan oleh pemilah s . jika simpul yang diperoleh merupakan kelas yang tidak homogen, prosedur yang sama diulangi sampai pohon klasifikasi menjadi suatu konfigurasi dan memenuhi:

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1) \quad (2.3)$$

b. Penentuan Simpul Terminal

Suatu simpul t akan menjadi simpul terminal atau tidak, akan dipilah kembali apabila pada simpul t tidak terdapat penurunan keheterogenan secara berarti atau adanya batasan minimum sebesar n seperti hanya terdapat satu pengamatan pada tiap simpul anak. Menurut Breiman, et al (1993), pengembangan pohon akan berhenti apabila pada simpul terdapat pengamatan berjumlah kurang dari atau sama dengan 5 ($n \leq 5$). Selain itu, proses pembentukan pohon juga akan berhenti apabila sudah mencapai batasan jumlah level yang telah ditentukan atau tingkat kedalaman (*depth*) dalam pohon maksimal.

c. Penandaan Label Kelas

Penentuan label kelas pada simpul terminal berdasarkan aturan jumlah terbanyak, yaitu jika

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.4)$$

dengan:

$p(j|t)$: proporsi kelas j pada simpul t

$N_j(t)$: jumlah pengamatan kelas j pada *terminal node* t

$N(t)$: jumlah total pengamatan pada *terminal node* t

Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t yang paling kecil sebesar $r(t) = 1 - \max_j p(j|t)$.

Proses pembentukan pohon klasifikasi berhenti ketika terdapat hanya satu pengamatan dalam tiap-tiap simpul anak atau adanya batasan minimum n , semua pengamatan dalam tiap simpul anak identik, dan adanya batasan jumlah level/kedalaman pohon maksimal. Setelah pembentukan pohon maksimal, tahap selanjutnya adalah pemangkasan pohon untuk mencegah terbentuknya pohon klasifikasi yang berukuran besar dan kompleks.

2.1.2 Pemangkasan Pohon Klasifikasi

Pohon yang dibentuk dengan aturan pemilah dan kriteria *goodness of split* berukuran sangat besar karena penghentian pohon berdasarkan banyaknya amatan pada simpul terminal atau besarnya tingkat kehomogenan. Jika semakin banyak pemilahan yang dilakukan, maka dapat mengakibatkan kecilnya tingkat kesalahan prediksi, akan tetapi akibatnya pohon klasifikasi yang dibentuk berukuran besar. Ukuran pohon yang besar dapat dapat memunculkan adanya *overfitting*, akan tetapi apabila pengamatan pohon dibatasi dengan ketepatan batas tertentu, maka dapat terjadi kasus *underfitting*. Oleh karena itu, untuk mendapatkan pohon yang layak, maka perlu dilakukan pemangkasan *pruning* yaitu suatu penilaian ukuran pohon tanpa mengorbankan akurasi yang berarti. Pemangkasan ini dilakukan dengan melakukan pengurangan simpul pohon sehingga dicapai ukuran pohon yang layak dan tidak terlalu melebar. Menurut Breiman dkk (1993), ukuran pohon yang layak dapat dilakukan dengan pemangkasan pohon dengan ukuran *cost complexity minimum*.

$$R_{\alpha}(T) = R(T) + \alpha |\tilde{T}| \quad (2.5)$$

dengan:

$R(T)$ = *Resubstitution Estimate* (proporsi kesalahan pada pohon T)

α = kompleksitas parameter (*complexity parameter*)

$|\tilde{T}|$ = ukuran banyaknya simpul terminal pohon T

$R_{\alpha}(t)$ merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan *cost penalty* bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. *Cost complexity pruning* menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan $R_{\alpha}(t)$ pada seluruh pohon bagian atau untuk setiap nilai α . Selanjutnya, dilakukan pencarian pohon bagian $T(\alpha) < T_{max}$ yang meminimumkan $R_{\alpha}(t)$ yaitu

$$R_{\alpha}(T(\alpha)) = \min_{T < T_{max}} R_{\alpha}(T) \quad (2.6)$$

Pemangkasan pohon dimulai dengan mengambil t_R dan t_L dari T_{max} yang dihasilkan dari simpul induk t . jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan $R(t) = R(t_R) + R(t_L)$, maka simpul anak t_R dan t_L dipangkas. Sehingga

diperoleh pohon T_i yang memenuhi kriteria $R(T_i) = R(T_{max})$. Jika $R(T)$ digunakan sebagai kriteria penentuan pohon optimal maka akan cenderung akan dipilih pohon terbesar T_i . sebab semakin besar pohon, maka semakin kecil nilai $R(T)$ nya.

2.1.3 Penentuan Pohon Klasifikasi Optimal

Pohon klasifikasi yang berukuran besar akan memberikan nilai *cost complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penduga pengganti yang cukup kecil. Apabila $R(T)$ dipilih sebagai penduga terbaik, maka cenderung akan dipilih pohon yang besar, sebab pohon yang semakin besar akan membuat nilai $R(T)$ semakin kecil. $R(T)$ atau *Resubstitution Estimate*/penduga pengganti merupakan proporsi amatan yang mengalami kesalahan pengklasifikasian.

Penduga pengganti ini sering digunakan apabila pengamatan yang ada tidak cukup besar. Pengamatan dalam L dibagi secara random menjadi V bagian yang saling lepas dengan ukuran kurang lebih sama besar untuk setiap kelas. Pohon $T^{(v)}$ dibentuk dari sampel *learning* ke- v dengan $v=1,2,\dots,V$. dimisalkan $d^{(v)}(x)$ adalah hasil pengklasifikasian, maka penduga sampel uji untuk $R(T_i^{(v)})$ adalah sebagai berikut.

$$R(T_i^{(v)}) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_v}^N X(d^{(v)}(x_n) \neq j_n) \quad (2.7)$$

Dengan $N_v \cong N/V$ adalah jumlah pengamatan dalam L_v .

Selanjutnya dilakukan prosedur yang sama dengan menggunakan semua pengamatan dalam L untuk membentuk deret pohon T_i . Penduga *cross validation v-fold* untuk $T_i^{(v)}$ adalah.

$$R^{cv}(T_i) = \frac{1}{V} \sum_{v=1}^V R^{cv}(T_i^{(v)}) \quad (2.8)$$

Pohon klasifikasi yang optimum dipilih T^* dengan $R^{cv}(T^*) = \min_i R^{cv}(T_i)$

2.1.4 Ilustrasi CART

Pembentukan pohon klasifikasi CART dapat diilustrasikan pada contoh berikut dengan melibatkan variabel respon Y berkategori dua dan dua variabel prediktor yang masing – masing berupa data kotegorik dan kontinu. Variabel yang digunakan dalam ilustrasi CART dijelaskan pada Tabel 2.1.

Tabel 2.1 Variabel Ilustrasi CART

Variabel	Nama Variabel	Kategori	Skala
Y	Status	0 : Tidak rekuren 1 : Rekuren	Nominal
X ₁	Usia	-	Rasio
X ₂	Jenis Terapi	0 : Tranfusi PRC 1 : Kemoterapi 2 : Operasi	Nominal

Ilustrasi CART dilakukan dengan menggunakan data bangkitan sebanyak 20 sampel pengamatan. Data sampel yang digunakan untuk ilustrasi pembentukan pohon klasifikasi CART ditampilkan pada Tabel 2.2 berikut.

Tabel 2.2 Data Sampel untuk Ilustrasi CART

Data sampel ke-	Y	X ₁	X ₂	Data sampel ke-	Y	X ₁	X ₂
1	2	57	3	11	1	45	1
2	1	57	2	12	2	60	2
3	1	60	2	13	1	71	2
4	2	45	1	14	2	35	3
5	1	35	2	15	1	60	2
6	1	45	1	16	2	71	2
7	1	57	1	17	2	35	1
8	2	60	2	18	1	57	1
9	2	35	3	19	2	45	2
10	1	57	2	20	1	45	3

Data sampel yang berjumlah 20 tersebut selanjutnya membentuk suatu simpul yang disebut sebagai simpul awal atau *parent node*. Langkah awal dalam analisis CART adalah

menghitung nilai keheterogenan simpul awal tersebut dengan menggunakan fungsi keheterogenan Indeks Gini yang telah didefinisikan pada persamaan (2.1). kemudian untuk memilah simpul awal menjadi dua simpul anak, maka dilakukan penentuan pemilah dari semua kemungkinan variabel dan nilai variabel (*threshold*). Terlebih dahulu dihitung nilai *goodness of split* (persamaan (2.2)) dari masing – masing kemungkinan *threshold* setiap variabel, kemudian dibandingkan dan dipilih *threshold* variabel yang memiliki nilai *goodness of split* paling tinggi (persamaan (2.3)). *Threshold* variabel terpilih tersebut selanjutnya dijadikan sebagai pemilah terbaik untuk simpul awal. Berikut diberikan contoh perhitungan manual dari proses pemilihan pemilah terbaik.

➤ Perhitungan Indeks Gini untuk variabel Y

Variabel Y mempunyai skala pengukuran nominal dengan dua kategori.

$N(t) = 20$ pengamatan

$N_i(t) = 11$ pengamatan $\rightarrow p(i|t) = 11/20 = 0,55$

$N_j(t) = 9$ pengamatan $\rightarrow p(j|t) = 9/20 = 0,45$

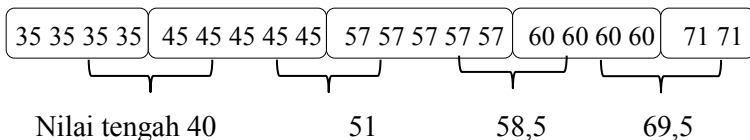
$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j$$

$$i(t) = p(1|t)p(2|t) + p(2|t)p(1|t)$$

$$i(t) = (0,55)(0,45) + (0,45)(0,55) = 0,495$$

➤ Penentuan *threshold* dan perhitungan *goodness of split* variabel X_1 (Usia pasien)

Variabel X_1 mempunyai skala pengukuran kontinu, oleh karena itu *threshold* variabel X_1 adalah nilai tengah (median) dari dua nilai amatan sampel berurutan yang berbeda dari variabel X_1 . Perhitungan nilai tengah untuk menentukan *threshold* ditunjukkan pada perhitungan berikut.



Tabel 2.3 Perhitungan *Goodness of Split* untuk *Threshold* sebesar 40

Simpul Kiri		Simpul Kanan	
$N_L = 4$	$p_L = 0,2$	$N_R = 16$	$p_R = 0,8$
$N(I t_L) = 1$	$p(I t_L) = 0,25$	$N(I t_R) = 10$	$p(I t_R) = 0,625$
$N(2 t_L) = 3$	$p(2 t_L) = 0,75$	$N(2 t_R) = 6$	$p(1 t_R) = 0,375$
$i(t_L) = (0,25)(0,75) + (0,75)(0,25)$ $= 0,375$		$i(t_R) = (0,625)(0,375) + (0,675)(0,325)$ $= 0,46875$	
<i>Goodness of Split</i> $\phi(s,t) = 0,495 - 0,2(0,375) - 0,8(0,46875) = 0,045$			

Tabel 2.4 Perhitungan *Goodness of Split* untuk *Threshold* sebesar 51

Simpul Kiri		Simpul Kanan	
$N_L = 9$	$p_L = 0,45$	$N_R = 11$	$p_R = 0,55$
$N(I t_L) = 4$	$p(I t_L) = 0,44$	$N(I t_R) = 7$	$p(I t_R) = 0,64$
$N(2 t_L) = 5$	$p(2 t_L) = 0,56$	$N(2 t_R) = 4$	$p(I t_R) = 0,36$
$i(t_L) = (0,44)(0,56) + (0,56)(0,44)$ $= 0,493$		$i(t_R) = (0,64)(0,36) + (0,36)(0,64)$ $= 0,461$	
<i>Goodness of Split</i> $\phi(s,t) = 0,495 - 0,45(0,493) - 0,55(0,461) = 0,0196$			

Tabel 2.5 Perhitungan *Goodness of Split* untuk *Threshold* sebesar 58,5

Simpul Kiri		Simpul Kanan	
$N_L = 14$	$p_L = 0,7$	$N_R = 6$	$p_R = 0,3$
$N(I t_L) = 8$	$p(I t_L) = 0,57$	$N(I t_R) = 3$	$p(I t_R) = 0,5$
$N(2 t_L) = 6$	$p(2 t_L) = 0,43$	$N(2 t_R) = 3$	$p(I t_R) = 0,5$
$i(t_L) = (0,57)(0,43) + (0,43)(0,57)$ $= 0,49$		$i(t_R) = (0,5)(0,5) + (0,5)(0,5)$ $= 0,5$	
<i>Goodness of Split</i> $\phi(s,t) = 0,495 - 0,7(0,49) - 0,3(0,5) = 0,002$			

Tabel 2.6 Perhitungan *Goodness of Split* untuk *Threshold* sebesar 69,5

Simpul Kiri		Simpul Kanan	
$N_L = 18$	$p_L = 0,9$	$N_R = 2$	$p_R = 0,1$
$N(I t_L) = 10$	$p(I t_L) = 0,56$	$N(I t_R) = 1$	$p(I t_R) = 0,5$
$N(2 t_L) = 8$	$p(2 t_L) = 0,44$	$N(2 t_R) = 1$	$p(I t_R) = 0,5$
$i(t_L) = (0,56)(0,44) + (0,44)(0,56)$ $= 0,493$		$i(t_L) = (0,5)(0,5) + (0,5)(0,5)$ $= 0,25$	
<i>Goodness of Split</i> $\phi(s,t) = 0,495 - 0,9(0,493) - 0,1(0,25) = 0,0013$			

- Penentuan *threshold* dan perhitungan *goodness of split* variabel X_2 (jenis terapi)

Variabel X_2 merupakan variabel kategorik dengan tiga kategori dan berskala nominal. Oleh karena itu, banyaknya kemungkinan *threshold* adalah:

$$2^{L-1} - 1 = 2^{3-1} - 1 = 3 \text{ kemungkinan pemilah}$$

Kemungkinan *threshold* yang pertama adalah kategori 1 sebagai simpul kiri dan kategori 2 & 3 sebagai simpul kanan. Hasil perhitungan *goodness of split* untuk kemungkinan *threshold* yang pertama ditunjukkan dalam Tabel 2.7.

Tabel 2.7 Perhitungan *Goodness of Split* untuk *Threshold* 1:2&3

Simpul Kiri		Simpul Kanan	
$N_L = 6$	$p_L = 0,3$	$N_R = 14$	$p_R = 0,7$
$N(1 t_L) = 4$	$p(1 t_L) = 0,67$	$N(1 t_R) = 7$	$p(1 t_R) = 0,5$
$N(2 t_L) = 2$	$p(2 t_L) = 0,33$	$N(2 t_R) = 7$	$p(1 t_R) = 0,5$
$i(t_L) = (0,67)(0,33)+(0,33)(0,67)$		$i(t_R) = (0,5)(0,5)+(0,5)(0,5)$	
$= 0,4422$		$= 0,5$	
<i>Goodness of Split</i> $\phi(s,t)= 0,495-0,3(0,4422)-0,7(0,5)=0,01234$			

Kemungkinan *threshold* berikutnya adalah kategori 2 sebagai simpul kiri dan kategori 1 & 3 sebagai simpul kanan. Hasil perhitungan *goodness of split* untuk kemungkinan *threshold* yang kedua ditunjukkan dalam Tabel 2.8.

Tabel 2.8 Perhitungan *Goodness of Split* untuk *Threshold* 2:1&3

Simpul Kiri		Simpul Kanan	
$N_L = 10$	$p_L = 0,5$	$N_R = 10$	$p_R = 0,5$
$N(1 t_L) = 6$	$p(1 t_L) = 0,6$	$N(1 t_R) = 5$	$p(1 t_R) = 0,5$
$N(2 t_L) = 4$	$p(2 t_L) = 0,4$	$N(2 t_R) = 5$	$p(1 t_R) = 0,5$
$i(t_L) = (0,6)(0,4)+(0,4)(0,6)$		$i(t_R) = (0,5)(0,5)+(0,5)(0,5)$	
$= 0,48$		$= 0,5$	
<i>Goodness of Split</i> $\phi(s,t)= 0,495-0,5(0,48)-0,5(0,5)=0,005$			

Kemungkinan *threshold* berikutnya adalah kategori 3 sebagai simpul kiri dan kategori 2 & 3 sebagai simpul kanan.

Hasil perhitungan *goodness of split* untuk kemungkinan *threshold* yang ketiga ditunjukkan dalam Tabel 2.9.

Tabel 2.9 Perhitungan *Goodness of Split* untuk *Threshold* 3:1&2

Simpul Kiri		Simpul Kanan	
$N_L = 4$	$p_L = 0,2$	$N_R = 16$	$p_R = 0,8$
$N(I t_L) = 1$	$p(I t_L) = 0,25$	$N(I t_R) = 10$	$p(I t_R) = 0,625$
$N(2 t_L) = 3$	$p(2 t_L) = 0,75$	$N(2 t_R) = 6$	$p(I t_R) = 0,375$
$i(t_L) = (0,25)(0,75)+(0,75)(0,25)$ $= 0,375$		$i(t_R) = (0,625)(0,375)+(0,625)(0,375)$ $= 0,4688$	
<i>Goodness of Split</i> $\phi(s,t) = 0,495-0,2(0,375)-0,8(0,4688)=0,04496$			

➤ Penentuan pemilah terbaik

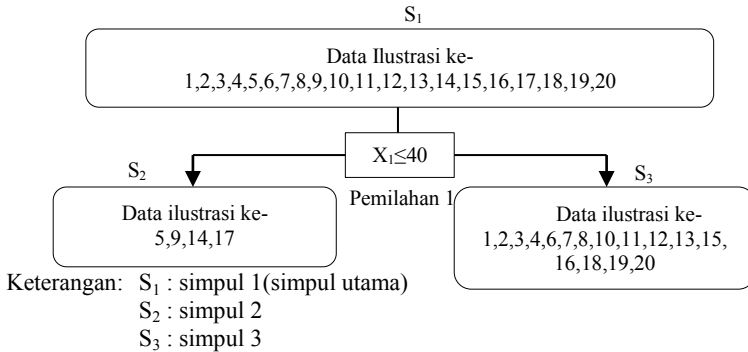
Tabel 2.10 Perbandingan *Goodness of Split* dari Setiap Kemungkinan Pemilah 1

Variabel	Threshold	<i>Goodness of Split</i> ($\phi(s,t)$)
X_1	40	0,045*
	51	0,0196
	58,5	0,002
	69,5	0,0013
X_2	1:2&3	0,01234
	2:1&3	0,005
	3:1&2	0,04496

**goodness of split* terbesar

Pemilah terbaik dipilih dari *threshold* variabel yang memiliki nilai *goodness of split* terbesar. Berdasarkan hasil perbandingan nilai *goodness of split* dari semua kemungkinan pemilah untuk variabel X_1 dan X_2 seperti yang tercantum dalam Tabel 2.12, dapat diketahui bahwa nilai *goodness of split* terbesar adalah pada variabel X_1 dengan *threshold* sebesar 40. Oleh karena itu, variabel utama dalam pemilah ini adalah variabel usia (X_1). Selanjutnya data sampel dipilah berdasarkan pemilah utama tersebut. Sehingga data sampel yang memiliki nilai X_1 kurang dari atau sama dengan *threshold* 40 masuk ke dalam simpul kiri dan sampel dengan nilai X_1 yang lebih besar dari 40 masuk ke dalam

simpul kanan. Hasil pohon yang terbentuk dari proses pemilahan ini ditunjukkan dalam Gambar 2.2.



Gambar 2.2 Pohon Klasifikasi Hasil Pemilahan Simpul Awal

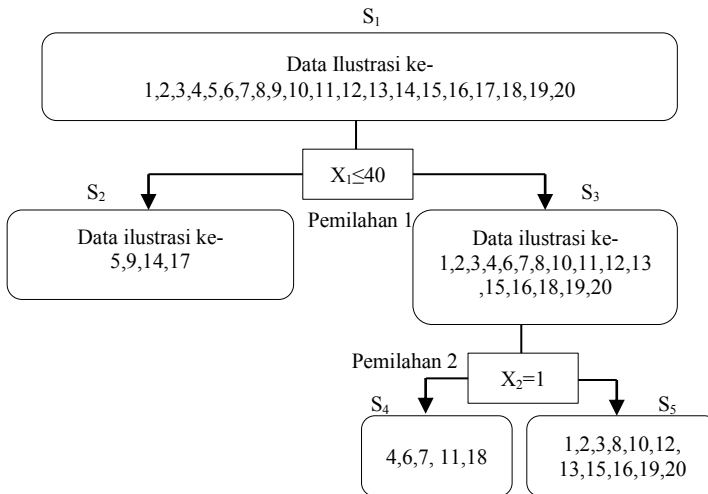
Pembentukan pohon dilakukan secara *recursive* atau terus menerus hingga didapatkan pohon yang maksimal. Oleh karena itu pohon yang dihasilkan pada Gambar 2.2 selanjutnya dilakukan pemilahan kembali hingga diperoleh simpul yang homogen. Pemilihan pemilah dari tiap simpul mengikuti langkah yang sama seperti pada proses pemilihan pemilah utama. Simpul kiri pada kedalaman 2 merupakan simpul terminal, sehingga tidak perlu dilakukan pemilahan kembali. Sedangkan simpul kanan dengan kedalaman 2 akan memilah sampel menjadi dua simpul anak lagi. Hasil perbandingan *goodness of split* setiap kemungkinan *threshold* dari sampel kanan ditunjukkan dalam Tabel 2.11.

Tabel 2.11 Perbandingan *Goodness of Split* dari Setiap Kemungkinan Pemilah 2

Variabel	Threshold	<i>Goodness of Split</i> ($\phi(s,t)$)
X ₁	51	0,0006
	58,5	0,0188
	69,5	0,0045
X ₂	1:2&3	0,0279*
	2:1&3	0,01245
	3:1&2	0,0045

* *goodness of split* terbesar

Tabel 2.11 menunjukkan bahwa dari semua kemungkinan pemilah, nilai *goodness of split* paling besar dihasilkan oleh pemilah X_2 (jenis pengobatan yang dijalani oleh pasien) pada kategori 1 (transfusi PRC) dengan nilai *goodness of split* sebesar 0.0279. Oleh karena itu, data sampel simpul kiri pada kedalaman 2 dipilah menjadi dua oleh pemilah X_2 yang berkategori 1. Data sampel yang memiliki nilai kategori X_2 sebesar 1 masuk ke simpul kiri dan sampel dengan nilai kategori X_2 sebesar 1 dan 2 masuk kedalam simpul kanan. Sehingga hasil pemilahan untuk simpul kanan dan kiri pada kedalaman 2 dapat digambarkan dalam sebuah pohon pada Gambar 2.3.



Keterangan: S_1 : simpul 1

S_2 : simpul 2, dan seterusnya.

Gambar 2.3 Pohon Klasifikasi Hasil Pemilahan 2

Berdasarkan gambar yang dihasilkan oleh Gambar 2.3, selanjutnya dilakukan pemilahan kembali hingga diperoleh simpul yang homogen. Simpul kiri pada kedalaman 3 merupakan simpul terminal sehingga tidak perlu dilakukan pemilahan lagi. Sedangkan simpul kanan pada kedalaman 3 akan memilah data sampel menjadi dua simpul anak lagi. Hasil perbandingan

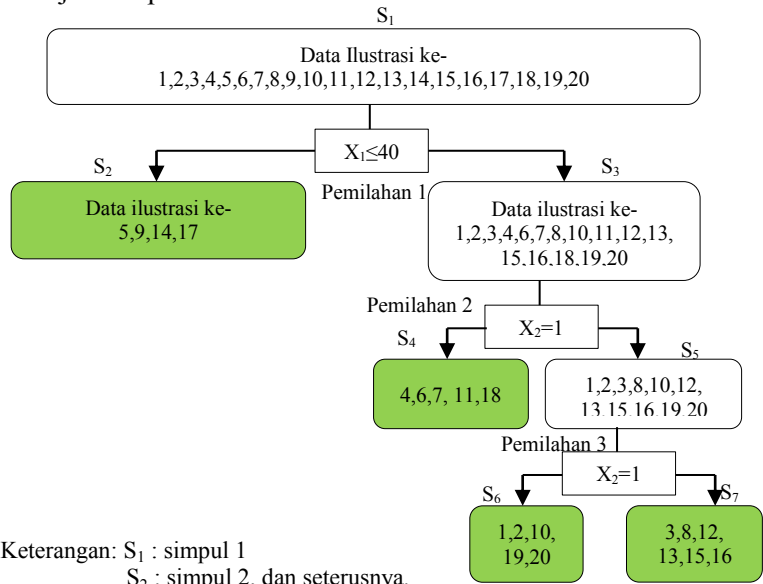
goodness of split setiap kemungkinan *threshold* dari simpul kanan ditunjukkan pada Tabel 2.12.

Tabel 2.12 Perbandingan *Goodness of Split* dari Setiap Kemungkinan Pemilah 3

Variabel	Threshold	Goodness of Split ($\phi(s,t)$)
X_1	51	0,00097
	58,5	0,00499*
	69,5	0,00097
X_2	2:3	0,00097

* *goodness of split* terbesar

Tabel 2.12 menunjukkan bahwa dari semua kemungkinan pemilah, nilai *goodness of split* paling besar dihasilkan oleh pemilah X_1 (usia) dengan nilai *goodness of split* sebesar 0.00499. oleh karena itu, data sampel yang memiliki nilai X_1 kurang dari atau sama dengan *threshold* 58.5 masuk kedalam simpul kiri dan sampel dengan nilai X_1 lebih dari 58.5 masuk kedalam simpul kanan. Hasil pohon yang terbentuk dari proses pemilahan ini ditunjukkan pada Gambar 2.4.



Gambar 2.4 Pohon Klasifikasi Hasil Pemilahan 3

Dimisalkan kriteria penghentian pembentukan pohon atau dengan kata lain pohon dikatakan homogen apabila data sampel ilustrasi pada simpul berjumlah enam pengamatan, maka pohon yang terbentuk pada Gambar 2.4 dikatakan sebagai pohon maksimal dan simpul yang paling akhir terbentuk dikatakan sebagai *terminal nodes* (ditunjukkan oleh kotak yang berwarna hijau). Sehingga dalam kasus ini diperoleh simpul terminal (*terminal nodes*) sebanyak empat simpul.

➤ **Penandaan Label Kelas**

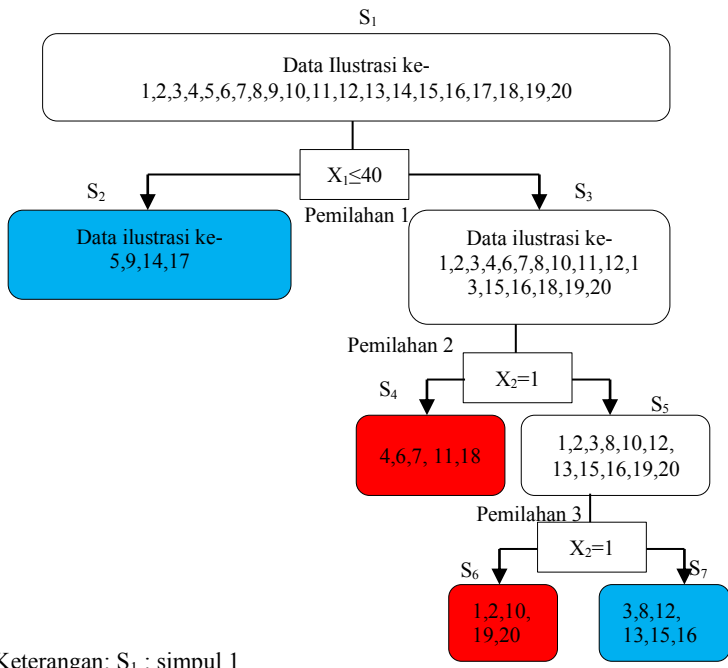
Langkah selanjutnya adalah pelabelan kelas pada setiap *terminal nodes*. Pemberian label kelas yaitu label kelas 1 untuk pasien yang tidak rekuren dan label kelas 2 untuk pasien rekuren, ditentukan berdasarkan persamaan (2.4) dengan mencari nilai peluang kelas $j=1$ yang maksimum diantara kedua simpul anak.

Tabel 2.13 Perhitungan Peluang Klasifikasi Simpul

Simpul	Terminal Nodes	$N(I t)$	$N(t)$	$p(I t)$	Nama Label Kelas
S_2	1	1	4	0,25	2
S_4	2	4	5	0,80	1
S_6	3	3	5	0,60	1
S_7	4	3	6	0,50	2

Berdasarkan Tabel 2.13 diketahui bahwa hasil pemilahan simpul S_1 menjadi simpul S_2 dan S_3 diperoleh nilai $p(j=I|t)$ untuk S_2 sebesar 0,25 dan untuk simpul S_3 masih dilanjutkan ke pemilahan selanjutnya. Karena nilai simpul S_2 memiliki nilai 0,25, maka simpul S_2 diberi label kelas 2. Begitu juga dengan simpul S_4 yang memiliki nilai $p(j=I|t)$ sebesar 0,80, simpul ini dilabeli kelas 1 karena simpul S_3 dipilah menjadi simpul S_4 dan S_5 tetapi S_5 dilanjutkan ke pemilahan selanjutnya. Sedangkan simpul S_5 dilakukan pemilahan menjadi simpul S_6 dan S_7 diperoleh nilai $p(j=I|t)$ untuk S_6 sebesar 0,60 dan nilai $p(j=I|t)$ untuk S_7 sebesar 0,50, karena nilai peluang untuk simpul S_6 lebih besar dari simpul S_7 maka simpul S_6 dilabeli kelas 1 dan simpul S_7 dilabeli kelas 2. Pada Gambar 2.5, simpul terminal dengan label kelas 1 (pasien

yang tidak rekuren) ditandai dengan simpul yang berwarna merah dan label kelas 2 (pasien yang rekuren) ditandai dengan simpul yang berwarna biru.



Gambar 4.5 Pohon Klasifikasi Setelah Dilakukan Pelabelan Kelas

Apabila peneliti menganggap pohon yang terbentuk terlalu kompleks, maka dapat dilakukan proses pemangkasan pohon (*pruning*) dengan mengikuti persamaan (2.5) dan selanjutnya dipilih pohon yang paling optimal berdasarkan kriteria yang ditentukan peneliti.

2.2 Ukuran Ketepatan Klasifikasi

Ketepatan klasifikasi digunakan untuk mengetahui apakah data diklasifikasikan dengan benar atau tidak (Agresti, 2002).

Ketepatan klasifikasi merupakan suatu evaluasi untuk melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi. Beberapa cara yang umum digunakan untuk mengukur ketepatan klasifikasi adalah melalui perhitungan *Apparent Error Rate* (APER), *sensitivity*, *specificity*, *total accuracy rate* (1-APER). APER merupakan proporsi observasi yang diprediksi secara tidak benar (ukuran kesalahan klasifikasi total). *Sensitivity* merupakan ukuran ketepatan dari kejadian yang diinginkan. *Specificity* merupakan ukuran yang menyatakan persentase kejadian yang tidak diinginkan. *Total accuracy rate* merupakan proporsi observasi yang diprediksi secara benar (ukuran kesalahan klasifikasi total) (Johnson dan Wichern, 1992).

Tabel untuk menghitung ketepatan klasifikasi ditunjukkan pada tabel 2.14.

Tabel 2.14 *Crosstab* Ketepatan Klasikasi

Observasi Y	Prediksi Y		Total
	1	2	
1	n_{11}	n_{12}	$N_{1.}$
2	n_{21}	n_{22}	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	N

Keterangan:

- n_{11} : Jumlah subjek dari variabel Y kategori 1 yang tepat diprediksikan sebagai variabel Y kategori 1
- n_{12} : Jumlah subjek dari variabel Y kategori 1 yang salah diprediksikan sebagai variabel Y kategori 2
- n_{21} : Jumlah subjek dari variabel Y kategori 2 yang salah diprediksikan sebagai variabel Y kategori 1
- n_{22} : Jumlah subjek dari variabel Y kategori 2 yang tepat diprediksikan sebagai variabel Y kategori 2
- $N_{1.}$: Jumlah observasi dari variabel Y kategori 1
- $N_{2.}$: Jumlah observasi dari variabel Y kategori 2
- $N_{.1}$: Jumlah prediksi dari variabel Y kategori 1
- $N_{.2}$: Jumlah prediksi dari variabel Y kategori 2
- N : Jumlah total observasi/prediksi

$$APER = \frac{\text{Jumlah prediksi salah}}{\text{Jumlah total prediksi}} = \frac{n_{12} + n_{21}}{N} \quad (2.9)$$

$$Sensitivity = \frac{n_{11}}{N_1} \quad (2.10)$$

$$Specificity = \frac{n_{22}}{N_2} \quad (2.11)$$

$$Total\ accuracy\ rate = 1 - APER = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah total prediksi}} = \frac{n_{11} + n_{22}}{N} \quad (2.12)$$

2.3 Cross Validation

Cross validation adalah metode yang sering digunakan untuk evaluasi model. *Cross Validation* merupakan salah satu teknik untuk menilai/ memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Pembuatan model biasanya bertujuan untuk melakukan prediksi maupun klasifikasi terhadap suatu data baru yang boleh jadi belum pernah muncul di dalam dataset. Data yang digunakan dalam proses pembangunan model disebut data *training*, sedangkan data yang akan digunakan untuk memvalidasi model disebut sebagai data *testing*.

Salah satu metode *cross validation* yang populer adalah *K-Fold Cross Validation*. Dalam teknik ini dataset dibagi menjadi sejumlah K -buah partisi secara acak. Kemudian dilakukan sejumlah K -kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke- K sebagai data *testing* dan memanfaatkan sisa partisi lainnya sebagai data *learning*.

Untuk mendapatkan nilai akurasi ataupun ukuran penilaian lainnya dari hasil eksperimen yang kita lakukan, dapat diambil nilai rata-rata dari seluruh eksperimen tersebut. Metode ini tidak lagi mempermasalahkan bagaimana data dibagi menjadi dua set karena semua *subset* akan menjadi data *training* maupun *testing* sehingga variansi dari hasil estimasi berkurang. Namun, metode ini membutuhkan waktu komputasi yang lebih lama karena algoritma *cross validation* harus diulang dari awal sebanyak k kali (Schneider, 1997).

Sebagai gambaran, jika kita melakukan *5-Fold Cross-Validation* maka desain data eksperimennya sebagai berikut.

Dataset:

K1 K2 K3 K4 K5

Data Eksperimennya ditunjukkan pada Tabel 2.15.

Tabel 2.15 Data Eksperimen untuk *5-Fold Cross Validation*

Percobaan Ke-	Data <i>Learning</i>	Data <i>Testing</i>
1	K2,K3,K4,K5	K1
2	K1,K3,K4,K5	K2
3	K1,K2,K4,K5	K3
4	K1,K2,K3,K5	K4
5	K1,K2,K3,K4	K5

2.4 Gambaran Mengenai Kanker Serviks

Kanker leher rahim (serviks) merupakan kanker pembunuh nomer dua di dunia setelah kanker payudara. Menurut departemen kesehatan RI tahun 2007, penyakit kanker leher rahim saat ini menempati urutan pertama daftar kanker yang diderita kaum wanita Indonesia. Saat ini ada sekitar 100 kasus per 100 ribu penduduk atau 200 ribu kasus setiap tahunnya. Kanker Serviks yang sudah masuk ke stadium lanjut sering menyebabkan kematian dalam jangka waktu relatif cepat.

Kanker serviks menyerang daerah leher rahim atau serviks yang disebabkan infeksi virus HPV yang tidak sembuh dalam waktu lama. Jika kekebalan tubuh berkurang, maka infeksi HPV akan mengganas dan dapat menyebabkan terjadinya kanker serviks. Gejalanya tidak terlalu kelihatan pada stadium dini, itulah sebabnya kanker serviks yang dimulai dari infeksi HPV dianggap sebagai "*The Silent Killer*". Perjalanan dari infeksi virus menjadi kanker membutuhkan waktu cukup lama, sekitar 10-20 tahun. Proses ini seringkali tidak disadari hingga kemudian sampai pada tahap pra kanker tanpa gejala (Meilinda, 2014).

Penyebab paling utama kanker serviks adalah HPV yang mempunyai diameter 55 μm dan virus ini ditularkan secara seksual. HPV memiliki kapsul isohedral yang telanjang dengan 72 kapsomer, serta mengandung DNA *circular double stranded*

dengan panjang kira-kira 8000 pasang basa. Penyebab banyak kematian pada kaum wanita adalah virus HPV tipe 16 dan 18. Virus ini sangat mudah berpindah dan menyebar, tidak hanya melalui cairan, tapi juga bisa berpindah melalui sentuhan kulit. Penggunaan wc umum yang sudah terkena virus HPV juga dapat menjangkit seseorang yang menggunakannya jika tidak membersihkannya dengan baik (Yamato, Fen, Kobuchi, Nasu, Yamada, Nishihara, et al, 2006).

Kanker serviks dapat terjadi jika infeksi HPV tidak sembuh dalam waktu yang lama. Apalagi dengan sistem imun atau kekebalan tubuh yang rendah, infeksi akan ganas dan menyebabkan sel kanker. Virus ini dapat menyebar melalui sentuhan, misalnya ada virus HPV di tangan, lalu menyentuh daerah genital, maka daerah serviks dapat terinfeksi.

2.4.1 Faktor Penyebab Kanker Serviks

Menurut Diananda (2009), Terdapat sejumlah faktor resiko atau penyebab kanker serviks, yaitu sebagai berikut.

1. Hubungan seksual di usia yang terlalu muda, berganti-ganti partner seks, atau berhubungan seks dengan pria yang sering berganti pasangan. Virus HPV dapat menular melalui hubungan seksual.
2. Wanita berusia diatas 40 tahun lebih rentan terkena kanker serviks. Semakin tua umur seseorang, maka semakin besar resiko terkena kanker serviks.
3. Memiliki terlalu banyak anak (lebih dari 5). Pada saat wanita melahirkan secara alami, janin akan melewati serviks dan akan menimbulkan trauma pada serviks, yang dapat memicu aktifnya sel kanker. Semakin sering janin melewati serviks, maka semakin sering trauma terjadi dan semakin tinggi resiko kanker serviks.
4. Daya tahan tubuh yang lemah, kurangnya konsumsi vitamin E, vitamin C dan asam folat. Kebiasaan merokok juga menambah resiko kanker serviks.

5. Pemakaian pembalut wanita yang mengandung bahan dioksin (bahan pemutih yang dipakai untuk memutihkan pembalut hasil daur ulang dari barang bekas).
6. Keputihan yang berlangsung terus menerus dan tidak diobati. Terdapat dua macam keputihan, yaitu keputihan normal dan tidak normal. Keputihan normal, lendir berwarna bening, tidak bau dan tidak gatal. Jika salah satu dari ketiga syarat tersebut tidak terpenuhi, maka termasuk keputihan tidak normal.
7. Memasukkan atau membersihkan genital dengan air yang tidak bersih. Air yang kotor mengandung banyak kuman dan bakteri.

2.4.2 Gejala dan Pengobatan Kanker Serviks

Pada stadium awal, gejala kanker serviks tidak terlalu kentara. Dibutuhkan waktu sekitar 10-20 tahun dari infeksi hingga menjadi tumor. Menurut Dalimartha (2004), Gejala-gejala kanker serviks adalah sebagai berikut.

1. Keluar darah berlebihan ketika menstruasi
2. Mengeluarkan sedikit darah ketika berhubungan badan
3. Terasa sakit ketika berhubungan seksual
4. Keputihan yang tidak normal
5. Pada stadium lanjut, gejalanya adalah sakit punggung atau tidak bisa berdiri tegak, salah satu paha bengkak, sakit di otot bagian paha, berat badan naik-turun, tidak dapat buang air kecil, bocornya urin/air seni dari vagina, tulang yang rapuh dan nyeri panggul, serta pendarahan spontan setelah *menopause*.

Menurut Alifah (2014), Terdapat sejumlah metode untuk mendeteksi kanker serviks, yaitu sebagai berikut.

1. IVA (Inspeksi Visual dengan Asam Asetat) merupakan metode pemeriksaan dengan mengoles serviks atau leher rahim dengan asam asetat. Kemudian diamati apakah ada kelainan seperti area berwarna putih. Jika tidak ada perubahan warna, maka dapat dianggap tidak ada infeksi pada serviks. Pemeriksaan ini dilakukan hanya untuk deteksi dini. Jika

terlihat tanda yang mencurigakan, maka metode deteksi lainnya yang lebih lanjut harus dilakukan.

2. *Pap Smear*, metode tes *Pap smear* yang umum dilakukan dengan menggunakan pengerik atau sikat untuk mengambil sedikit sampel sel-sel serviks atau leher rahim. Kemudian sel-sel tersebut akan dianalisa di laboratorium. Tes itu dapat menyingkapkan apakah ada infeksi, radang, atau sel-sel abnormal. Menurut laporan sedunia, dengan secara teratur melakukan tes *Pap smear* telah mengurangi jumlah kematian akibat kanker serviks.
3. *Thin Prep* merupakan metode berbasis cairan yang lebih akurat dari *pap smear*, karena *pap smear* hanya mengambil sebagian dari sel-sel di serviks atau leher rahim, maka *thin prep* akan memeriksa seluruh bagian serviks atau leher rahim. Tentu hasilnya akan jauh lebih akurat dan tepat.
4. *Kolposkopi*, prosedur ini dilakukan jika hasil tes pada metode sebelumnya menunjukkan adanya infeksi. Prosedur kolposkopi dilakukan dengan menggunakan alat yang dilengkapi lensa pembesar untuk mengamati bagian yang terinfeksi. Tujuannya untuk menentukan apakah ada lesi atau jaringan yang tidak normal pada serviks atau leher rahim.

2.4.3 Stadium Kanker Serviks

Stadium kanker serviks menurut pada sistem klasifikasi dari FIGO (*Federation of Gyenaecologic and Obstetrics*) tahun 2000 dilihat berdasarkan lokasi tumor primer, ukuran besar tumor, dan adanya penyebaran keganasan (WHO, 2008; Pelsi, 2006) ditunjukkan pada Tabel 2.16. Stadium ini dibuat untuk mempermudah perencanaan terapi yang efektif dan optimal bagi pasien dan memperkirakan prognosis pasien.

Tabel 2.16 Stadium Kanker Serviks

Stadium FIGO (%)	Kategori TNM	5-tahun Survive
0	Tumor Utama tidak bisa diperiksa Tidak ada bukti tentang tumor utama Karsinoma prainvasif	
I	Karsinoma terbatas pada kandungan IA Karsinoma serviks berdasar pemeriksaan mikroskopis IA1 Invasi stroma dengan kedalaman $\leq 3,00$ mm dan invasi horizontal $\leq 7,00$ mm IA2 Invasi stroma $> 3,00$ mm dan $\leq 5,00$ dengan suatu invasi horizontal 7,00 atau lebih sedikit IB Tampak lesi secara klinis, terbatas pada serviks, atau lesi mikroskopis yang lebih besar dari 1A1/1A2 IB1 Lesi $< 4,00$ mm IB2 Lesi $> 4,00$ mm	90 - 95% 80 - 85%
II	Tumor invasif di luar kandungan, tapi tidak sampai dinding panggul atau sepertiga bawah vagina IIA Tanpa invasi ke parametrium IIB Dengan invasi ke parametrium	50 - 65% 40 - 50%
III	Tumor meluas ke dinding panggul atau melibatkan sepertiga bawah vagina dan/ menyebabkan hidronefrosis atau tidak berfungsi nya ginjal IIIA Tumor melibatkan sepertiga bawah vagina tanpa perluasan ke dinding panggul IIIB Tumor meluas ke dinding panggul dan/ menyebabkan hidronefrosis atau tidak berfungsi nya ginjal	25 - 30%
IV	Tumor meluas ke luas pelvis atau secara klinis melibatkan mukosa kandung kemih dan/ rektum IVA Tumor invasi ke mukosa kandung kemih atau rektum dan/ meluas di luar tulang panggul IVB Metastasis jauh	$< 5\%$

Sumber : FIGO (*Federation of Gyenaecologic and Obstetrics*) tahun 2000

2.4.4 Rekurensi Kanker Serviks

Rekurensi kanker terjadi apabila setelah pengobatan (radiasi/operasi) dan pasien dinyatakan remisi (hilangnya secara lengkap atau parsial dari tanda-tanda dan gejala penyakit sebagai respon terhadap pengobatan) kemudian kanker timbul kembali baik secara hispatologis maupun terlihat secara klinis (Moeloek, Nuranna, Wibowo,& Purbadi, 2006). Menurut Chamim (2006), kasus rekuren terjadi hampir 90% dalam dua tahun pertama paskaterapi. Pengertian kasus residif menurut Standar Pelayanan Medik Perkumpulan Obstetri dan Ginekologi Indonesia tahun 2006, dibagi menjadi beberapa macam berdasarkan jenis terapi yang telah diberikan.

1. Sembuh primer post radiasi

Apabila serviks ditutup oleh epitel normal atau obliterasi vagina tanpa adanya ulkus atau cairan yang keluar. Pada pemeriksaan rektovaginal, jika terdapat indurasi teraba licin, tidak berbenjol. Serviks besarnya tidak lebih dari 2,5 cm dan tidak ada metastasis jauh

2. Rekuren post radiasi

Apabila tumor tumbuh kembali di pelvis setelah serviks dan vagina dinyatakan sembuh.

3. Rekuren post operatif

Apabila ditemukan masa tumor post operatif dimana masa tumor sudah terangkat secara makroskopik dan tepi sayatan dinyatakan bebas secara histologik.

4. Kanker baru

Timbul lesi lokal setelah paling sedikit 10 tahun sesudah radiasi pertama

2.4.5 Faktor – faktor Penyebab Rekurensi Kanker Serviks

Penelitian Dunst, Kuhnt, Strauss, Krause, Pelz, Koelbl, H., et al. (2003) menyimpulkan bahwa anemia memiliki dampak yang signifikan terhadap tingkat kekambuhan. Penelitian Loka, Tsukuma, Ajiki, & Oshima (2005) menyimpulkan bahwa wanita dengan usia antara 30–54 tahun memiliki resiko terkena kanker serviks lebih tinggi dibandingkan dengan wanita usia 55-64 tahun

dan lebih dari 65 tahun. Penelitian yang dilakukan oleh Matsuura, Kawagoe, Toki, Tanaka, & Kashimura (2006) memberikan kesimpulan bahwa ada komplikasi setelah 5 tahun pengobatan kanker serviks. Dalam jurnal yang ditulis oleh Choi, Fuller, Thomas, & Wang (2008), jenis kanker dan stadium kanker juga menjadi faktor yang penting. Pasien kanker serviks stadium 1 cenderung lebih banyak dibandingkan dengan stadium lainnya. Sedangkan untuk jenis kanker serviks, pasien dengan penderita jenis kanker serviks karsinoma epidermoid lebih banyak dibandingkan dengan jenis kanker serviks lainnya. Berdasarkan penelitian Lee, Kim, Kim, Choi, Do, Song, et al (2013), terdapat 38 wanita berusia antara 51-65 tahun dari 300 wanita yang mengalami rekurensi kanker serviks. Lee, et al (2013) juga menyebutkan bahwa dari 38 kasus rekurensi yang terjadi, pasien yang menjalani pengobatan operasi dengan adjuvant CCRT (*Concurrent Chemo-Radiation Therapy*) cenderung lebih tinggi risikonya mengalami rekurensi kanker serviks dibandingkan dengan hanya menjalani operasi saja dan operasi dengan adjuvant RT (*Radiotherapy*). Sedangkan berdasarkan penelitian Novitasari (2014), stadium dan jenis kanker serviks berpengaruh terhadap rekurensi kanker serviks.

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Dalam penelitian ini, data yang digunakan merupakan data sekunder yang diperoleh dari data rekam medis pasien kanker serviks di RSUD dr. Soetomo Surabaya. Data yang digunakan pada penelitian ini merupakan data pasien kanker serviks tahun 2014 yang melakukan rawat inap dengan jumlah data sebanyak 810 pasien. Struktur datanya adalah sebagai berikut.

Tabel 3.1 Struktur Data

Data ke-	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	y ₁	x ₁₁	x ₂₁	x ₃₁	x ₄₁	x ₅₁	x ₆₁
2	y ₂	x ₁₂	x ₂₂	x ₃₂	x ₄₂	x ₅₂	x ₆₂
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
810	y ₈₁₀	x ₈₁₀	x ₈₁₀	x ₈₁₀	x ₈₁₀	x ₈₁₀	x ₈₁₀

3.2 Variabel Penelitian

Variabel penelitian yang digunakan pada penelitian ini meliputi variabel dependen (Y) dan variabel independen (X). variabel dependen yang digunakan dalam penelitian ini adalah status rekurensi kanker serviks, rekurensi yang dimaksudkan adalah kembalinya pasien ke RSUD dr. Soetomo Surabaya karena penyakit yang sama yaitu kanker serviks yang berskala nominal, terdiri dari.

Y = 1 jika tidak rekuren

Y = 2 jika rekuren

Sedangkan variabel prediktor yang digunakan pada penelitian ini merupakan faktor-faktor yang mempengaruhi rekurensi kanker serviks. Deskripsi lengkap dari variabel prediktor yang digunakan pada penelitian ini dapat dilihat pada tabel 3.2.

Tabel 3.2. Variabel Penelitian

Variabel	Nama Variabel	Kategori	Skala
X ₁	Usia Pasien	-	Rasio
X ₂	Status Anemia	1 : Tidak ada 2 : Ada	Nominal
X ₃	Stadium kanker serviks	0 : Stadium 0 1: Stadium I 2 : Stadium II 3 : Stadium III 4 : Stadium IV	Ordinal
X ₄	Jenis pengobatan yang dijalani pasien	1 : Transfusi PRC 2 : Kemoterapi 3 : Kemoterapi & Tranfusi PRC 4 : Operasi	Nominal
X ₅	Status Komplikasi	1 : Tidak ada 2 : Ada	Nominal
X ₆	Status Penyakit Penyerta	1 : Bukan Penyakit Penyerta 2: Penyakit Penyerta	Nominal

3.3 Metode Analisis Data

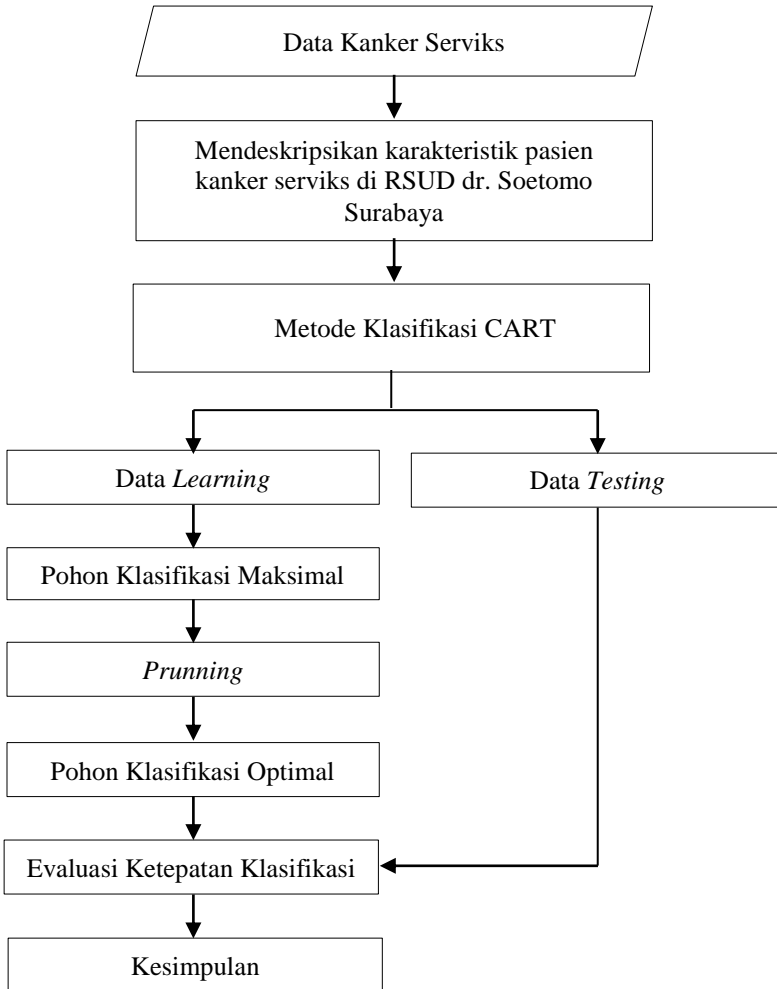
Metode analisis data yang digunakan pada penelitian ini adalah

1. Mendeskripsikan pasien kanker serviks dengan melakukan analisis statistika deskriptif pada variabel-variabel yang diduga berpengaruh terhadap rekurensi kanker serviks, yaitu usia pasien (X₁), status anemia (X₂), stadium (X₃), jenis pengobatan (X₄), status komplikasi(X₅) dan status penyakit penyerta (X₆).
2. Membagi data menjadi dua bagian dengan metode *5-fold Cross Validation*. Jadi data penelitian dibagi menjadi *k-subset*, kemudian diulang sebanyak *k* kali. Setiap pengulangan, salah satu dari *k subset* digunakan untuk data *testing* dan sisanya, *k-1 subset*, digunakan untuk membentuk model (data *training*). Selanjutnya membandingkan hasil ketepatan klasifikasi dari setiap *fold* untuk data *learning* dan *testing*. Kombinasi data yang dipilih adalah kombinasi data yang

memiliki tingkat akurasi tertinggi dan jumlah terminal *nodes* yang tidak terlalu banyak.

3. Melakukan metode klasifikasi CART melalui tahapan berikut.
 - a. Membentuk pohon klasifikasi maksimal menggunakan data *learning* dengan langkah – langkah berikut.
 1. Melakukan pemilihan pemilah dari variabel prediktor berdasarkan aturan pemilahan Indeks Gini, kemudian hasil pemilahan dievaluasi dengan menggunakan kriteria *Goodness of Split*.
 2. Menentukan terminal *nodes* dengan cara menghentikan pembentukan pohon hingga mencapai batasan minimum pengamatan dalam terminal *node* sebanyak 1 pengamatan.
 3. Melakukan penandaan label kelas pada terminal *nodes* berdasarkan aturan jumlah terbanyak dari tiap kelas pada variabel respon.
 - b. Memangkas pohon (*pruning*) berdasarkan kriteria *cost complexity minimum*.
 - c. Memilih pohon klasifikasi yang optimal dengan *test sample estimates*.
 - d. Menghitung ketepatan klasifikasi pohon hasil bentukan dengan menggunakan data *learning* dan mengevaluasinya dengan menjalankan data *testing* pada pohon klasifikasi yang terbentuk. Ukuran ketepatan klasifikasi yang digunakan adalah APER (*Apparent Error Rate*), *sensitivity*, *specificity* dan *total accuracy rate* (1-APER).

Tahapan metode analisis data selengkapnya disajikan pada Gambar 3.1.



Gambar 3.1 Diagram Alir Analisis Data

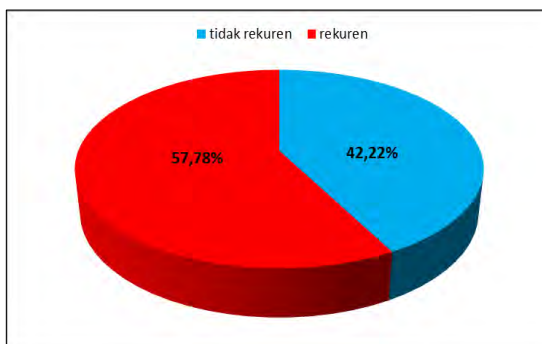
BAB IV

ANALISIS DAN PEMBAHASAN

Pada bab ini akan dilakukan analisis data pada faktor-faktor yang mempengaruhi rekurensi kanker serviks di RSUD dr. Soetomo Surabaya untuk menjawab permasalahan dan tujuan dari penelitian yang telah dikemukakan pada bab sebelumnya.

4.1 Deskripsi Karakteristik Pasien

Rekurensi kanker serviks yang dimaksudkan pada penelitian ini adalah kembalinya pasien ke RSUD dr. Soetomo Surabaya karena penyakit yang sama, yaitu kanker serviks. Pada penelitian ini digunakan enam variabel yang diduga berpengaruh terhadap rekurensi kanker serviks diantaranya usia, status anemia, stadium, jenis pengobatan yang diikuti pasien, status komplikasi dan status penyakit penyerta. Untuk mendapatkan gambaran secara menyeluruh mengenai pasien kanker serviks di RSUD dr. Soetomo Surabaya, maka perlu dilakukan deskripsi secara statistik dari 6 variabel yang berpengaruh terhadap rekurensi kanker serviks. Deskripsi mengenai status rekurensi kanker serviks di RSUD dr. Soetomo Surabaya disajikan dalam Gambar 4.1 berikut.



Gambar 4.1 Deskripsi Status Rekurensi Kanker Serviks

Gambar 4.1 menjelaskan mengenai status rekurensi kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014. Dari 810 pasien yang menjadi subjek penelitian, dapat diketahui bahwa pada tahun 2014 pasien kanker serviks di RSUD dr. Soetomo Surabaya lebih banyak yang rekuren, yaitu sebesar 57,78 persen atau sebanyak 468 pasien. Sedangkan yang tidak rekuren sebesar 42,22 persen atau sebanyak 342 pasien.

4.1.1 Karakteristik Pasien Berdasarkan Usia Pasien

Penyakit kanker serviks menyerang semua orang tanpa melihat usia. Usia yang dimaksudkan disini adalah usia pasien ketika datang ke RSUD dr. Soetomo Surabaya. Deskripsi status rekurensi kanker serviks di RSUD dr. Soetomo Surabaya berdasarkan usia pasien ditunjukkan pada tabel 4.1.

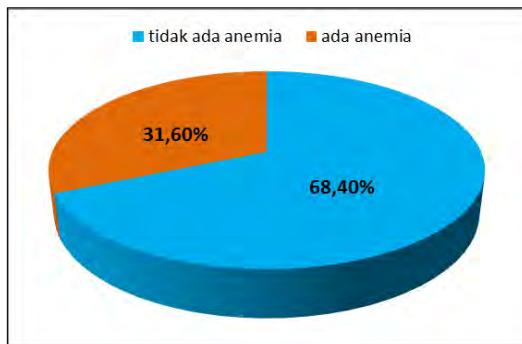
Tabel 4.1 Deskripsi Usia Pasien Berdasarkan Status Rekurensi (Tahun)

		<i>Mean</i>	<i>Varians</i>	<i>Minimum</i>	<i>Median</i>	<i>Maximum</i>
Y	Tidak Rekuren	50,620	86,008	27	50	79
	Rekuren	49,346	63,777	27	49	72

Tabel 4.1 menunjukkan bahwa pada pasien yang tidak rekuren, rata-rata pasien kanker serviks berusia 50,62 tahun dengan usia minimal 27 tahun dan usia maksimal 79 tahun, sedangkan median usia pasien 50 tahun dengan varians sebesar 86,008. Sedangkan pada pasien yang mengalami rekuren kanker serviks, rata-rata usia pasien kanker serviks adalah 49,346 tahun dengan usia minimum 27 tahun dan usia maksimum 72 tahun, median usia pasien adalah 49 tahun dengan varians sebesar 63,777.

4.1.2 Karakteristik Pasien Berdasarkan Faktor Status Anemia

Status anemia merupakan salah satu faktor yang mempengaruhi rekurensi kanker serviks. Deskripsi status anemia pada pasien rekurensi kanker serviks di RSUD dr. Soetomo tahun 2014 ditunjukkan pada Gambar 4.2.



Gambar 4.2 Deskripsi Status Anemia

Gambar 4.2 menunjukkan bahwa jumlah pasien yang tidak mengalami anemia lebih banyak dibandingkan pasien yang mengalami anemia. Jumlah pasien yang tidak mengalami anemia sebesar 68,40 persen atau sebanyak 554 pasien, sedangkan sisanya sebanyak 31,60 persen mengalami anemia.

Deskripsi status anemia jika dilihat berdasarkan status rekurensi kanker serviks disajikan pada Tabel 4.2.

Tabel 4.2 Deskripsi Status Anemia Berdasarkan Status Rekurensi

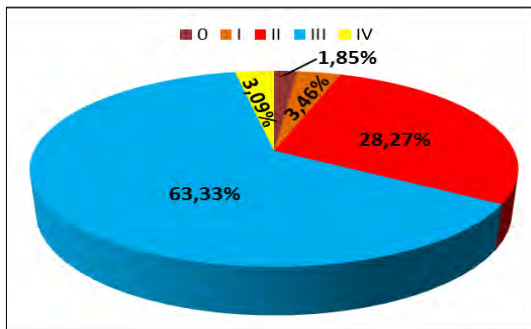
		Status Rekurensi		Total
		Tidak Rekuren	Rekuren	
Status Anemia	Tidak ada anemia	179 (52,3%)	375 (80,1%)	554 (68,4%)
	Ada anemia	163 (47,7%)	93 (19,9%)	256 (31,6%)
	Total	342 (100%)	256 (100%)	810 (100%)

Tabel 4.2 menjelaskan mengenai status anemia berdasarkan pada status rekurensi. Sebanyak 342 pasien tidak rekuren, 52,3 persen atau sebanyak 179 pasien diantaranya tidak mengalami anemia, sedangkan sisanya sebanyak 163 pasien atau sebesar 47,7 persen mengalami anemia. pada pasien yang rekuren, dari 256 pasien yang rekuren terdapat sebanyak 80,1 persen atau sebesar 375 pasien tidak mengalami anemia, sisanya sebanyak 19,9 persen mengalami anemia. sehingga dapat disimpulkan bahwa

mayoritas pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014 tidak mengalami anemia, baik pada pasien yang tidak rekuren maupun yang rekuren kanker serviks.

4.1.3 Karakteristik Pasien Berdasarkan Faktor Stadium Kanker Serviks

Faktor selanjutnya yang merupakan faktor yang mempengaruhi rekurensi kanker serviks adalah stadium kanker serviks. Stadium kanker serviks terdiri dari stadium 0, stadium I, stadium II, stadium III, dan stadium IV. Stadium menunjukkan tingkat keparahan kanker serviks, semakin tinggi stadiumnya, maka kanker serviksnya semakin parah. Deskripsi dari stadium kanker serviks yang diderita oleh pasien kanker serviks di RSUD dr. Soetomo ditunjukkan pada Gambar 4.3.



Gambar 4.3 Deskripsi Stadium Kanker Serviks

Gambar 4.3 menunjukkan bahwa stadium yang paling banyak diderita oleh pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014 adalah stadium III, yang ditunjukkan oleh persentase sebesar 63,33 persen atau sebanyak 513 pasien. Sedangkan stadium terbanyak kedua yang diderita oleh pasien adalah stadium II, yaitu sebesar 28,27 persen atau sebanyak 229 pasien. Sedangkan sisanya sebesar 3,09 persen menderita stadium IV, sebesar 3,46 persen menderita stadium I dan sebesar 1,85 menderita stadium 0.

Deskripsi stadium kanker serviks jika dilihat berdasarkan status rekurensi kanker serviks disajikan pada Tabel 4.3.

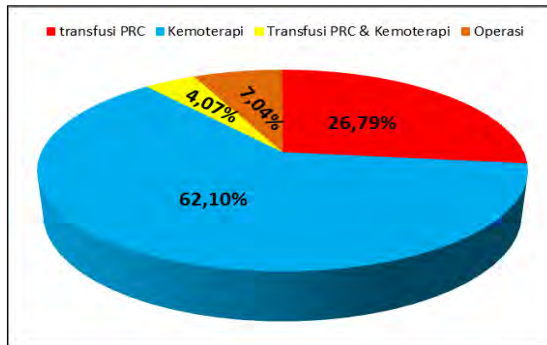
Tabel 4.3 Deskripsi Stadium Berdasarkan Status Rekurensi

	Status Rekurensi		Total
	Tidak rekuren	Rekuren	
0	14 (4,1%)	1 (0,2%)	15 (1,9%)
I	18 (5,3%)	10 (2,1%)	28 (3,5%)
Stadium II	83 (24,3%)	146 (31,2%)	229 (28,3%)
III	210 (61,4%)	303 (64,7%)	513 (63,3%)
IV	17 (5%)	8 (1,7%)	25 (3,1%)
Total	342 (110%)	468 (100%)	810 (100%)

Tabel 4.3 menjelaskan bahwa dari sebanyak 342 pasien yang tidak rekuren, sebesar 61,4 persen atau sebanyak 210 pasien menderita stadium III, stadium terbanyak kedua yang diderita oleh pasien adalah stadium II sebanyak 83 pasien atau sebesar 24,3 persen. Hal yang sama juga terjadi pada pasien yang tidak rekuren, dari sebanyak 468 pasien yang tidak rekuren, sebesar 63,3 persen atau sebanyak 513 pasien menderita kanker serviks stadium III dan untuk stadium II sebesar 31,2 persen atau sebanyak 146 pasien. Sehingga dapat disimpulkan bahwa stadium kanker serviks yang paling banyak diderita oleh pasien adalah stadium III, baik pada pasien yang rekuren dan pasien yang rekuren.

4.1.4 Karakteristik Pasien Berdasarkan Jenis Pengobatan yang Dijalani oleh Pasien

Variabel jenis pengobatan yang dijalani oleh pasien merupakan faktor yang juga berpengaruh terhadap rekurensi kanker serviks. Jenis pengobatan yang dijalani oleh pasien diantaranya adalah transfusi PRC, kemoterapi, kombinasi transfusi PRC & kemoterapi, dan operasi. Deskripsi dari stadium pasien kanker serviks ditunjukkan oleh Gambar 4.4.



Gambar 4.4 Deskripsi Jenis Pengobatan yang Dijalani Pasien

Gambar 4.4 menunjukkan bahwa jenis pengobatan yang paling banyak dijalani oleh pasien adalah kemoterapi, yang ditunjukkan oleh persentase sebesar 62,10 persen atau sebanyak 503 pasien. Selain itu, pasien yang menjalani pengobatan transfusi PRC sebesar 26,79 persen atau sebanyak 26,79%. Sedangkan pasien yang menjalani pengobatan operasi sebesar 7,04% atau sebanyak 57 pasien dan pasien yang menjalani kombinasi pengobatan tranfusi PRC dan kemoterapi sebesar 4,07 persen atau sebanyak 33 pasien.

Deskripsi jenis pengobatan yang dijalani oleh pasien jika dilihat berdasarkan status rekurensi ditunjukkan pada Tabel 4.4.

Tabel 4.4 Deskripsi Jenis Pengobatan yang Dijalani Pasien Berdasarkan Status Rekurensi

		Status Rekurensi		Total
		Tidak rekuren	Rekuren	
Jenis Pengobatan	Transfusi PRC	140 (40,9%)	77(16,5%)	217(26,8%)
	Kemoterapi	141 (41,2%)	362(77,4%)	503(62,1%)
	Transfusi PRC & Kemoterapi	16 (4,7%)	17(3,6%)	33(4,1%)
	Operasi	45 (13,2%)	12(2,6%)	57(7%)
Total		342 (100%)	468(100%)	810(100%)

Tabel 4.4 menjelaskan bahwa pada pasien yang tidak rekuren, jenis pengobatan yang paling banyak dijalani oleh pasien adalah kemoterapi yaitu sebesar 41,2 persen atau sebanyak 141 pasien, selain itu jenis pengobatan yang juga banyak dijalani oleh pasien adalah transfusi PRC yaitu sebesar 40,9 persen atau sebanyak 140 pasien. Sedangkan untuk pasien yang rekuren, jenis pengobatan yang paling banyak dijalani oleh pasien adalah kemoterapi, yaitu sebesar 77,4 persen atau sebanyak 362 pasien. Sehingga dapat disimpulkan bahwa pada pasien yang rekuren dan pasien yang tidak rekuren, jenis pengobatan yang paling banyak dijalani oleh pasien adalah kemoterapi.

4.1.5 Karakteristik Pasien Berdasarkan Status Komplikasi

Faktor selanjutnya yang berpengaruh terhadap rekurensi kanker serviks adalah status komplikasi. Ada atau tidaknya komplikasi berpengaruh terhadap rekurensi kanker serviks. Deskripsi status komplikasi pada pasien kanker serviks di RSUD dr. Soetomo ditunjukkan pada Gambar 4.5.



Gambar 4.5 Deskripsi Status Komplikasi

Gambar 4.5 menunjukkan bahwa mayoritas pasien kanker serviks di RSUD dr. Soetomo tidak mengalami komplikasi, dengan persentase sebesar 77,90 persen atau sebanyak 631 pasien. Sedangkan jumlah pasien yang mengalami komplikasi adalah 179 pasien atau sebanyak 22,10 persen.

Deskripsi status komplikasi pasien jika dilihat berdasarkan status rekurensi ditunjukkan pada Tabel 4.5.

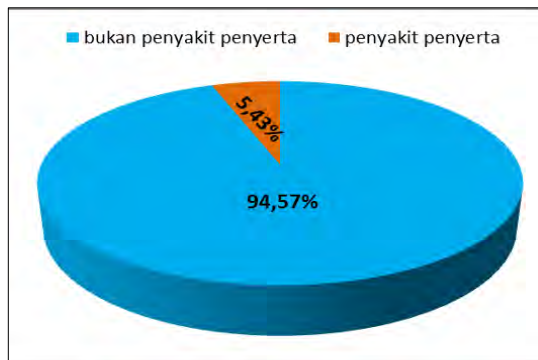
Tabel 4.5 Deskripsi Status Komplikasi Berdasarkan Status Rekurensi

		Status Rekurensi		Total
		Tidak rekuren	Rekuren	
Status Komplikasi	Tidak ada	232	399	631
	komplikasi	(67,8%)	(85,3%)	(77,9%)
	Ada komplikasi	110	69	179
		(32,2%)	(14,7%)	(22,1%)
Total		342	468	810
		(100%)	(100%)	(100%)

Tabel 4.5 menjelaskan bahwa pada pasien yang tidak rekuren, mayoritas pasien tidak mengalami komplikasi, yang ditunjukkan oleh persentase sebesar 67,8 persen atau sebanyak 232 pasien. Begitu juga pada pasien yang rekuren, mayoritas pasien tidak mengalami komplikasi yang ditunjukkan oleh persentase sebesar 85,3 persen atau sebanyak 399 pasien. Sehingga dapat disimpulkan bahwa mayoritas pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014 tidak mengalami komplikasi.

4.1.6 Karakteristik Pasien Berdasarkan Status Penyakit Penyerta

Selain variabel-variabel yang telah dijelaskan sebelumnya, variabel yang juga berpengaruh terhadap rekurensi kanker serviks adalah variabel status penyakit penyerta. Definisi dari penyakit penyerta yang dimaksudkan adalah kanker serviks yang diderita oleh pasien merupakan penyakit utama atau bukan (penyakit penyerta). Deskripsi dari status penyakit penyerta ditunjukkan pada Gambar 4.6.



Gambar 4.6 Deskripsi Status Penyakit Penyerta

Gambar 4.6 menunjukkan bahwa mayoritas penyakit kanker serviks yang diderita oleh pasien kanker serviks di RSUD dr. Soetomo Surabaya bukan merupakan penyakit penyerta, melainkan sebagai penyakit utama yang ditunjukkan oleh persentase sebesar 94,57 persen atau sebanyak 766 pasien. Sedangkan banyaknya pasien yang penyakit kanker serviks yang dideritanya merupakan penyakit penyerta hanya sebanyak 44 pasien atau sebesar 5,43 persen.

Deskripsi status penyakit penyerta jika dilihat berdasarkan status rekurensi ditunjukkan pada Tabel 4.6.

Tabel 4.6 Deskripsi Status Penyakit Penyerta Berdasarkan Status Rekurensi

		Status Rekurensi		Total
		Tidak rekuren	Rekuren	
Status Penyakit Penyerta	Bukan penyakit	315	451	766
	penyerta	(92,1%)	(96,4%)	(94,6%)
	Penyakit penyerta	27	17	44
		(7,9%)	(3,6%)	(5,4%)
Total		342	468	810
		(100%)	(100%)	(100%)

Tabel 4.6 menunjukkan bahwa pada pasien yang tidak rekuren, mayoritas penyakit kanker serviks yang diderita oleh pasien bukan merupakan penyakit penyerta, ditunjukkan oleh persentase sebesar 92,1 persen atau sebanyak 315 pasien. Hal yang sama juga terjadi pada pasien yang rekuren kanker serviks, mayoritas penyakit kanker serviks yang diderita oleh pasien bukan merupakan penyakit penyerta, ditunjukkan oleh persentase sebesar 96,4 persen atau sebanyak 451 pasien. Sehingga dapat disimpulkan bahwa mayoritas penyakit kanker serviks yang diderita oleh pasien bukan merupakan penyakit penyerta.

4.2 Analisis CART untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya

Data pengamatan yang digunakan untuk mengklasifikasi rekurensi pasien kanker serviks sebesar 810 pengamatan. Data tersebut dibagi menjadi dua bagian yaitu data *learning* dan data *testing*. Pembagian data menjadi dua bagian ini bertujuan untuk mendapatkan suatu model yang optimal. Pembagian data *learning* dan *testing* ini menggunakan *cross validation*. Pada penelitian ini, data pasien kanker serviks akan terlebih dahulu dibagi menjadi set *learning* dan *testing* menggunakan *5-fold Cross Validation* sebelum menghitung performansi model klasifikasi. Nantinya, setiap *fold* yang terbentuk akan menjadi data *learning* maupun *testing* secara bergiliran sehingga hasil evaluasi performansi model akan lebih baik karena memiliki variasi estimasi yang lebih kecil. Data *learning* berguna untuk mengestimasi model, sementara data *testing* digunakan untuk menguji model yang dibentuk oleh data *learning*.

Analisis CART diawali dengan langkah pembentukan pohon klasifikasi yang maksimal, yang kemudian dilakukan pemangkasan sampai diperoleh pohon klasifikasi yang optimal. Berikut penjelasan secara lebih mendalam untuk masing-masing tahapan analisis klasifikasi CART dengan menggunakan *5-fold Cross Validation*.

4.2.1 Pembentukan Pohon Klasifikasi Maksimal

Langkah pertama yang dilakukan untuk membentuk pohon klasifikasi adalah dengan menentukan variabel pemilah dan nilai variabel (*threshold*). Variabel pemilah dan *threshold* dipilih dari beberapa kemungkinan pemilah dari masing-masing variabel. Perhitungan banyaknya kemungkinan pemilah ditampilkan pada Tabel 4.7.

Tabel 4.7 Perhitungan Pemilah pada Setiap Variabel

Variabel	Nama Variabel	Skala Data	Banyaknya Kategori (nilai amatan sampel)	Kemungkinan Pemilah
X_1	Usia	Rasio	51	$51 - 1 = 50$ pemilah
X_2	Status anemia	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_3	Stadium	Ordinal	4	$4 - 1 = 3$ pemilah
X_4	Jenis pengobatan	Nominal	4	$2^{4-1} - 1 = 7$ pemilah
X_5	Status komplikasi	Nominal	2	$2^{2-1} - 1 = 1$ pemilah
X_6	Status penyakit penyerta	Nominal	2	$2^{2-1} - 1 = 1$ pemilah

Dari berbagai kemungkinan pemilah dari setiap variabel, langkah selanjutnya adalah menghitung Indeks Gini yang merupakan ukuran keheterogenan simpul. Cara kerja Indeks Gini adalah melakukan pemilahan simpul dengan berfokus pada masing-masing simpul kanan dan kiri. Selanjutnya, hasil perhitungan Indeks Gini kemudian digunakan untuk menentukan *goodness of split* dari masing-masing pemilah. Pemilah yang terpilih adalah variabel pemilah dan nilai variabel (*threshold*) yang memiliki nilai *goodness of split* tertinggi. Pemilah yang terpilih merupakan variabel yang terpenting dalam mengklasifikasikan data pengamatan. Besarnya kontribusi variabel sebagai pemilah baik pemilah utama maupun pengganti pada

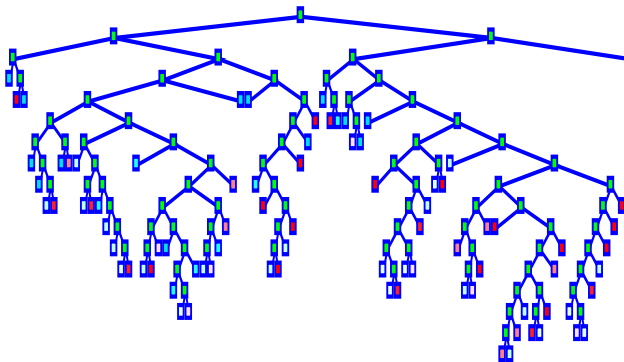
pohon klasifikasi maksimal yang terbentuk ditunjukkan melalui suatu angka skor ditampilkan pada Tabel 4.8.

Tabel 4.8 Skor Variabel Terpenting dari Pohon Klasifikasi Maksimal

Variabel	Skor Variabel	
X ₄	100.00	
X ₂	71.51	
X ₅	31.27	
X ₃	23.80	
X ₁	13.44	
X ₆	11.99	

Tabel 4.8 menunjukkan bahwa semua variabel menjadi pembangun pohon klasifikasi. Akan tetapi, berdasarkan skor yang dihasilkan diketahui bahwa variabel yang terpenting dan menjadi pemilah utama adalah jenis pengobatan yang dialami pasien (X₄) di RSUD dr. Soetomo Surabaya karena memiliki skor yang paling tinggi, yaitu sebesar 100. Selain itu terdapat beberapa variabel yang juga berpengaruh besar dalam proses pemilahan yaitu status anemia (X₂) sebesar 71.51 dan status komplikasi (X₅) sebesar 31.27. Skor variabel untuk variabel pembentuk lain selengkapnya dapat dilihat pada Tabel 4.8.

Konstruksi pohon klasifikasi maksimal berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks ditunjukkan pada Gambar 4.7.

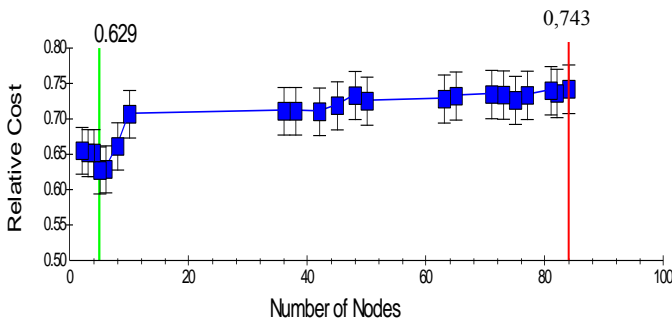


Gambar 4.7 Konstruksi Pohon Klasifikasi Maksimal

Hasil penyekatan rekursif secara biner dari data pengamatan yang digunakan akan menghasilkan pohon klasifikasi yang berukuran relatif besar dan tingkat kedalaman (*depth*) yang tinggi. Pohon klasifikasi tersebut disebut pohon klasifikasi maksimal. Gambar 4.7 menunjukkan hasil pohon klasifikasi maksimal yang berukuran besar yaitu dengan jumlah terminal *node* sebanyak 84 *node* dan dengan tingkat kedalaman sebesar 17.

4.2.2 Pemangkasan Pohon Klasifikasi Maksimal (*Prunning*)

Pohon klasifikasi maksimal yang dihasilkan kemudian dilakukan pemangkasan pohon secara iteratif berdasarkan kriteria tertentu yang telah dilakukan peneliti. Setiap hasil pemangkasan memiliki nilai *relative cost* tertentu, sehingga kemudian dipilih hasil pemangkasan dengan nilai *relative cost* minimum.



Gambar 4.8 Plot *Relative Cost* dan Banyaknya *Terminal Nodes*

Gambar 4.8 diperoleh dari hasil pemangkasan dengan menggunakan metode *test sample estimate*. Algoritma dari *test sample estimate* ini adalah dengan membagi data menjadi dua bagian yaitu L_1 untuk data *learning* dan L_2 untuk data *testing*. Data *learning* berguna untuk membentuk pohon T , sedangkan data *testing* berguna untuk pendugaan atau mengestimasi *misclassification rate error* atau disebut juga *resubstitution estimate* (penduga pengganti). Pada Gambar 4.8, pohon klasifikasi optimal ditunjukkan oleh garis berwarna hijau dan

pohon klasifikasi maksimal ditunjukkan oleh garis berwarna merah.

Urutan pembentukan pohon klasifikasi (*tree sequence*) ditunjukkan pada Tabel 4.9.

Tabel 4.9 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*)

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Test Set Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity Parameter</i>
1	84	$0,743 \pm 0,034$	0,454	0,000
14	38	$0,712 \pm 0,034$	0,511	$0,9799 \times 10^{-3}$
15	36	$0,712 \pm 0,034$	0,515	0,001
16	10	$0,708 \pm 0,034$	0,572	0,001
17	8	$0,663 \pm 0,033$	0,577	0,001
18	6	$0,630 \pm 0,033$	0,587	0,002
19*	5	$0,629 \pm 0,033$	0,593	0,003
20	4	$0,653 \pm 0,033$	0,603	0,005
21	3	$0,653 \pm 0,033$	0,618	0,008
22	2	$0,657 \pm 0,033$	0,639	0,010
23	1	$1,000 \pm 0,000$	1,000	0,181

*Pohon Klasifikasi Optimal

Tabel 4.9 menunjukkan bahwa pohon maksimal yang terbentuk terdiri dari 84 *terminal nodes* dan menghasilkan nilai *complexity parameter* sebesar 0,000, *test set relative cost* sebesar $0,743 \pm 0,034$ dan *resubstitution relative cost* paling kecil yaitu sebesar 0,454.

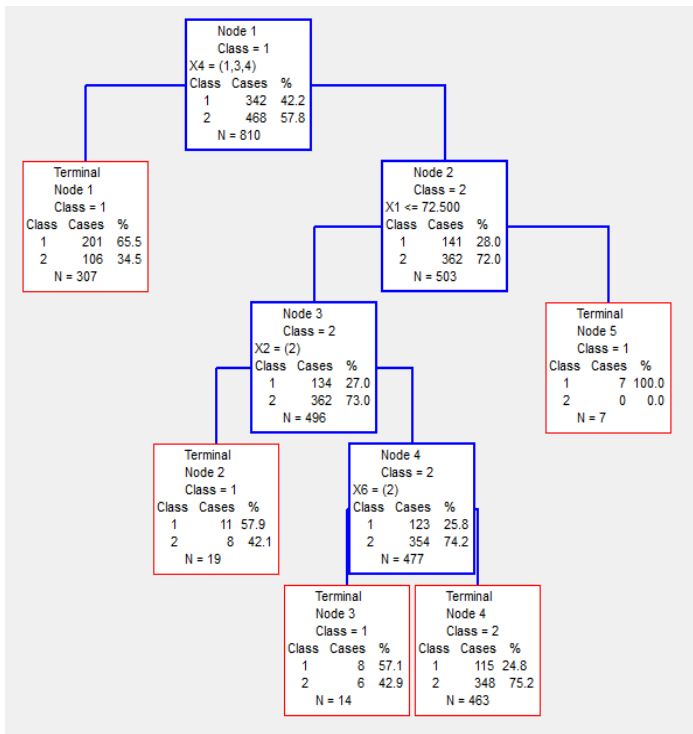
4.2.3 Pemilihan Pohon Klasifikasi Optimal

Hasil pemangkasan pohon secara iteratif yang diperoleh pada Gambar 4.8 selanjutnya digunakan untuk memilih pohon klasifikasi yang optimal. Pohon klasifikasi yang terpilih ditunjukkan pada Tabel 4.9 dengan ukuran pohon sebesar 19, *terminal nodes* berjumlah 5 simpul, nilai *complexity parameter* sebesar 0,003, *test set relative cost* paling minimum yaitu sebesar $0,629 \pm 0,033$ dan *resubstitution relative cost* sebesar 0,593.

Konstruksi pohon klasifikasi optimal dengan tingkat kedalaman pohon (*depth*) sebesar 5 disajikan pada Gambar 4.9.

Setiap simpul pada pohon klasifikasi optimal yang terbentuk seperti pada Gambar 4.9 telah dilakukan pelabelan kelas menjadi dua yaitu berwarna biru apabila suatu kelas diklasifikasikan sebagai pasien yang tidak rekuren kanker serviks, dan berwarna merah apabila suatu simpul diklasifikasikan sebagai pasien yang mengalami rekuren kanker serviks. Jika semakin kuat warna yang dihasilkan oleh masing-masing simpul, maka hal ini menandakan bahwa perbedaan proporsi antara kedua status rekurensi kanker serviks semakin besar.

Konstruksi pohon klasifikasi optimal berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks disajikan pada Gambar 4.9.



Gambar 4.9 Pohon Klasifikasi Optimal

Gambar 4.9 menunjukkan bahwa *terminal nodes* yang dihasilkan pada pohon klasifikasi maksimal sebanyak 5 *terminal nodes*. Variabel yang paling menentukan klasifikasi pasien kanker serviks berdasarkan rekurensi kanker serviks adalah variabel jenis pengobatan yang dijalani oleh pasien (X_4) di RSUD dr. Soetomo. Selain itu, terdapat tiga variabel lain yang juga berpengaruh dalam pengklasifikasian pasien kanker serviks yaitu variabel usia (X_1), status anemia (X_2) dan status penyakit penyerta (X_3). Skor dari keempat variabel tersebut disajikan pada Tabel 4.10.

Tabel 4.10 Skor Variabel Terpenting dari Pohon Klasifikasi Optimal

Variabel	Skor Variabel	
X_4	100,00	
X_1	13,12	
X_2	7,64	
X_6	5,80	

Tabel 4.10 menunjukkan bahwa jenis pengobatan yang jalani oleh pasien kanker serviks (X_4) di RSUD dr. Soetomo merupakan variabel pemilah yang utama dan paling menentukan klasifikasi pasien kanker serviks berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks dengan skor sebesar 100.00. Selain itu, terdapat tiga variabel lain yang juga memiliki kontribusi dalam pembentukan pohon klasifikasi diantaranya adalah usia (X_1) dengan skor sebesar 13,12, status anemia (X_2) sebesar 7,64 dan status penyakit penyerta (X_6) berkontribusi sebesar 5,80.

Proses Pemilahan diawali dengan memilah data sebanyak 810 pasien berdasarkan jenis pengobatan yang dijalani oleh pasien dan nilai *threshold* 1, 3 dan 4 (transfusi PRC, kombinasi kemoterapi dan transfusi PRC dan operasi). Kemudian data dibagi menjadi dua bagian yaitu simpul kiri (simpul 2) terdiri dari 307 pasien yang menjalani pengobatan transfusi PRC, kemoterapi & transfusi PRC dan operasi, dan selanjutnya dilabeli sebagai pasien yang tidak mengalami rekuren kanker serviks karena proporsi terbesar pada simpul tersebut adalah pasien yang tidak rekuren.

Sedangkan simpul kanan (simpul 3) terdiri dari 538 pasien yang menjalani pengobatan kemoterapi dengan label kelas sebagai pasien yang mengalami rekurensi. Selanjutnya, setiap simpul dilakukan pemilahan kembali karena masih memiliki tingkat keheterogenan yang tinggi. Pemilahan dilakukan sampai memperoleh *terminal node* sebanyak 5 simpul. Proses pemilahan sampai diperoleh *terminal node*.

Secara keseluruhan, berdasarkan Gambar 4.9 dapat diketahui bahwa terdapat 4 simpul terminal yang diklasifikasikan sebagai pasien yang tidak rekuren kanker serviks dan sebanyak 1 simpul terminal yang diklasifikasikan sebagai pasien yang rekuren kanker serviks. Karakteristik simpul terminal yang mengklasifikasikan data menjadi salah satu dari kedua status rekurensi dijelaskan sebagai berikut.

1. Simpul terminal 1 terdiri dari 307 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang menjalani pengobatan transfusi PRC, kombinasi kemoterapi dan transfusi PRC, dan operasi.
2. Simpul terminal 2 terdiri dari 19 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang memiliki anemia, berusia kurang dari atau sama dengan 72,5 tahun dan menjalani pengobatan kemoterapi.
3. Simpul terminal 3 terdiri dari 14 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang penyakit kanker serviks yang dideritanya merupakan penyakit penyerta, tidak memiliki anemia, berusia kurang dari atau sama dengan 72,5 tahun dan menjalani pengobatan kemoterapi.
4. Simpul terminal 4 terdiri dari 463 pasien kanker serviks yang diprediksi sebagai pasien yang rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang penyakit kanker serviks yang dideritanya bukan

merupakan penyakit penyerta, tidak memiliki anemia, berusia kurang dari atau sama dengan 72,5 tahun dan menjalani pengobatan kemoterapi.

5. Simpul terminal 5 terdiri dari 7 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang berusia lebih dari 72,5 tahun dan menjalani pengobatan kemoterapi.

Berdasarkan penjelasan dari kelima simpul terminal, maka karakteristik pasien kanker serviks berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks adalah sebagai berikut.

Tabel 4.11 Karakteristik Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya

Tidak Rekuren	Rekuren
<ul style="list-style-type: none"> • Pasien yang menjalani pengobatan transfusi PRC, kemoterapi & transfusi PRC dan operasi • Pasien yang menjalani pengobatan kemoterapi yang berusia $\leq 72,5$ tahun dan ada anemia • Pasien yang menjalani pengobatan kemoterapi yang berusia $\leq 72,5$ tahun, tidak ada anemia dan kanker serviks yang dideritanya merupakan penyakit penyerta • Pasien yang menjalani pengobatan kemoterapi dan berusia $> 72,5$ tahun 	<ul style="list-style-type: none"> • Pasien yang menjalani pengobatan kemoterapi yang berusia $\leq 72,5$ tahun, tidak ada anemia dan kanker serviks yang dideritanya bukan penyakit penyerta

4.2.4 Ketepatan Klasifikasi Pohon CART

Pada dasarnya, model klasifikasi di bidang medis digunakan untuk mendeteksi ada tidaknya penyakit dalam diri seorang pasien sehingga performasi klasifikasi tidak hanya dilihat dari nilai *total accuracy rate* (1-APER), namun juga meliputi nilai *sensitivity* dan *specificity*. Kemampuan model untuk menunjukkan hasil positif terhadap pasien yang benar-benar

mengidap suatu penyakit diukur menggunakan *sensitivity*, sementara kemampuan model untuk menunjukkan hasil negatif pada pasien yang memang tidak menderita suatu penyakit diukur menggunakan *specificity*.

Hasil ketepatan klasifikasi untuk data *learning* dengan menggunakan *5-fold cross validation* untuk masing-masing *fold* disajikan pada Tabel 4.12

Tabel 4.12 Klasifikasi Data *Learning* dengan Menggunakan *5-fold CV*

		Prediksi		
		Tidak rekuren	Rekuren	
Observasi	Tidak rekuren	1	204	69
		2	198	75
		3	194	73
		4	202	72
		5	195	79
	Rekuren	1	81	294
		2	65	310
		3	72	309
		4	64	310
		5	71	303

Penghitungan akurasi, *sensitivity*, dan *specificity* untuk *fold 1* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{204 + 294}{204 + 69 + 81 + 294} \right) \times 100\% = 76,85\%$$

$$\text{Sensitivity} = \left(\frac{204}{204 + 69} \right) \times 100\% = 74,73\%$$

$$\text{Specificity} = \left(\frac{294}{81 + 294} \right) \times 100\% = 78,4\%$$

Penghitungan yang sama juga dilakukan terhadap *fold 2* sampai dengan *fold 5*.

Berdasarkan perhitungan akurasi, *sensitivity* dan *specificity* yang telah dilakukan sebelumnya, maka hasil ketepatan klasifikasi untuk data *learning* pada permasalahan klasifikasi pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014 untuk masing-masing *fold* disajikan dalam Tabel 4.13.

Tabel 4.13 Evaluasi Performansi pada Data *Learning*

<i>Fold</i>	Akurasi	<i>Sensitivity</i>	<i>Specificity</i>
1	76,85%	74,73%	78,40%
2	78,40%	72,53%	82,67%
3	77,62%	72,66%	81,10%
4	79,01%	73,72%	82,89%
5	76,85%	71,17%	81,02%
Rata-rata	77,75%	72,96%	81,22%

Tabel 4.13 menunjukkan bahwa model terbaik untuk klasifikasi menggunakan *5-fold cross validation* adalah *fold* 4 karena memiliki nilai akurasi yang paling tinggi yaitu sebesar 79,01 persen, nilai *sensitivity* sebesar 72,96 dan *specificity* sebesar 82,89 persen. Selain itu, rata-rata akurasi yang diperoleh sebesar 77,75 persen, rata-rata nilai sensitivitas dan spesifisitas masing-masing sebesar 72,96 persen dan 81,22 persen.

Sedangkan ketepatan klasifikasi untuk data prediksi ditunjukkan pada Tabel 4.14. Ketepatan klasifikasi ini untuk menguji apakah pohon yang dibentuk oleh data *learning* sudah baik/tidak.

Tabel 4.14 Ketepatan Klasifikasi Untuk Data *Prediksi*

Observasi	Prediksi		Total
	Tidak rekuren	Rekuren	
Tidak rekuren	222	120	342
Rekuren	130	338	468
Total	352	458	810

Perhitungan akurasi, *sensitivity* dan *specificity* untuk data prediksi adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{222 + 338}{810} \right) \times 100\% = 69,14\%$$

$$\text{Sensitivity} = \left(\frac{222}{342} \right) \times 100\% = 64,91\%$$

$$\text{Specificity} = \left(\frac{338}{468} \right) \times 100\% = 72,22\%$$

Dari perhitungan yang dilakukan, nilai akurasi untuk data *testing* sebesar 69,14 persen, maka dapat dikatakan bahwa pohon optimal yang terbentuk dari data *learning* sudah cukup baik dan sesuai jika digunakan untuk mengklasifikasikan data baru. Sedangkan nilai *sensitivity* dan *specificity* masing-masing adalah sebesar 64,91 persen dan 72,22 persen.

Halaman ini Sengaja Dikosongkan

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, maka dapat diperoleh kesimpulan sebagai berikut.

1. Karakteristik pasien kanker di RSUD dr. Soetomo Surabaya berdasarkan status rekurensi adalah sebagai berikut.
 - Berdasarkan status rekurensi kanker serviks, lebih banyak pasien yang rekuren kanker serviks.
 - Berdasarkan karakteristik umur, rata-rata usia pasien yang tidak rekuren kanker serviks adalah 50,62 tahun, sedangkan rata-rata usia pasien yang rekuren kanker serviks adalah 49,32 tahun.
 - Mayoritas pasien kanker serviks di RSUD dr. Soetomo Surabaya tidak mengalami anemia.
 - Stadium kanker serviks yang banyak diderita oleh pasien adalah stadium III, baik itu pada pasien yang tidak rekuren kanker serviks maupun pada pasien yang rekuren kanker serviks.
 - Jenis terapi yang paling banyak dijalani oleh pasien kanker serviks adalah kemoterapi, hal ini berlaku untuk pasien yang tidak rekuren kanker serviks maupun pada pasien yang rekuren kanker serviks.
 - Persentase pasien yang tidak mengalami komplikasi sama-sama tinggi pada pasien yang tidak rekuren kanker serviks maupun pasien yang rekuren kanker serviks.
 - Mayoritas penyakit kanker serviks yang diderita oleh pasien bukan merupakan penyakit penyerta, melainkan sebagai penyakit utama.
2. Klasifikasi pasien kanker serviks berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks dilakukan dengan menggunakan pembagian data *learning* dan data *testing* dengan metode *5-fold Cross Validation*. Jumlah *terminal nodes* yang diperoleh sebanyak 5. Hasil klasifikasi

menunjukkan bahwa variabel yang berpengaruh besar dalam pembentukan pohon dan rekurensi kanker serviks adalah variabel jenis pengobatan yang dijalani oleh pasien. Selain itu, terdapat tiga variabel lain yang juga berpengaruh dalam pembentukan pohon dan rekurensi kanker serviks yaitu variabel usia, status anemia dan status penyakit penyerta. Jumlah *terminal nodes* yang diperoleh sebanyak 5 terbagi menjadi dua yaitu sebanyak 4 *terminal nodes* diklasifikasikan sebagai pasien yang tidak rekuren kanker serviks, dan 1 *terminal nodes* diklasifikasikan sebagai pasien yang rekuren kanker serviks. Sehingga karakteristik pasien berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks adalah:

- Karakteristik pasien yang tidak rekuren adalah pasien yang menjalani pengobatan transfusi PRC, kemoterapi & transfusi PRC dan operasi; pasien yang menjalani pengobatan kemoterapi yang berusia $\leq 72,5$ tahun dan ada anemia; pasien yang menjalani pengobatan kemoterapi yang berusia $\leq 72,5$ tahun, tidak ada anemia dan kanker serviks yang dideritanya merupakan penyakit penyerta; dan pasien yang menjalani pengobatan kemoterapi dan berusia $> 72,5$ tahun
- Karakteristik pasien yang rekuren adalah pasien yang menjalani pengobatan kemoterapi yang berusia $\leq 72,5$ tahun, tidak ada anemia dan kanker serviks yang dideritanya bukan penyakit penyerta.

5.2 Saran

Berdasarkan analisis yang telah dilakukan dan hasil pembahasan yang didapatkan maka terdapat beberapa saran yang dapat diberikan, yaitu.

1. Hasil akurasi yang diperoleh kurang besar, sehingga untuk penelitian selanjutnya disarankan menggunakan metode klasifikasi yang lain, antara lain CHAID, SVM, dan metode

klasifikasi lainnya agar nilai akurasi yang diperoleh lebih besar.

2. Untuk penelitian selanjutnya, dapat dipertimbangkan penambahan faktor-faktor yang berpengaruh terhadap rekurensi kanker serviks, misalnya jumlah pasitas, kadar hemoglobi, dan faktor-faktor lain yang berpengaruh terhadap rekurensi kanker serviks.
3. Jumlah data yang digunakan penelitian juga ditambah/diperbanyak, misalnya mengambil data selama 3 tahun. Karena kemungkinan hasil yang diperoleh juga akan lebih informatif.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

- Adrijono. (2007). *Kanker Serviks*. Divisi Onkologi, Dept. Obstetri-Ginekologi FKUI
- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley and Sons, Inc., New York
- Alifah, H. (2014). *Cara Medeteksi Kanker Serviks*. Diperoleh dari: <http://cara.media/mendeteksi-kanker-serviks/>. Diakses tanggal 26 Maret 2015
- Almarogi, S. (2015). *Gejala dan Penyebab Kanker Serviks*. Diperoleh dari : <http://www.infosehat.id/gejala-penyebab-kanker-serviks/>. Diakses tanggal 26 Maret 2015
- Breiman L., Friedman J.H., Olshen R.A., & Stone C.J. (1993). *Classification And Regression Trees*. New York: Chapman And Hall
- Chamim. (2006). *Buku Acuan Nasional Onkologi*. In: M Farid Aziz, Adrijono, Abdul Bari Saifuddin, editors. Penentuan stadium klinik dan pembedahan kanker ginekologi. Jakarta : Yayasan Bina Pustaka Sarwono Prawirohadjo. p. 173 - 181
- Choi, M., Fuller, C. D., Thomas, C. R., Wang, S. J. (2008). Conditional Survival In Ovarian Cancer : Result From The SEER Dataset 1988 – 2001. *Journal Gynecologic Oncology*. 109. 203 – 209
- Dalimartha, S. (2004). *Deteksi Dini Kanker dan Simplicia Anti Kanker*. Jakarta : Penebar Swadaya Jakarta
- Dewi, C. C. (2013). *Pengaruh Terapi Menulis Terhadap Penurunan Tingkat Kecemasan Pasien Kanker Serviks Stadium Lanjut di POSA RSUD Dr. Soetomo Surabaya*. Skripsi Program Studi Pendidikan Bidan Universitas Airlangga
- Diananda. (2009). *Panduan Lengkap Mengernai Kanker*. Yogyakarta : Mirza Media Pustaka
- Depkes RI. 2007. Pencegahan Kanker Leher Rahim dan Kanker Payudara. Jakarta.

- Dunst, J., Kuhnt, T., Strauss, H., Krause, U., Pelz, T., Koelbl, H., et al. (2003). Anemia in Cervical Cancers : Impact on Survival, Pattern of Relapse, and Association with Hypoxia and Angiogenesis. *International Journal of Radiation*. 56. 778 – 787
- International Union Against Cancer (UICC). (2009). *Jika Tidak Dikendalikan 26 Juta Orang di Dunia Menderita Kanker*. Diperoleh dari: <http://www.depkes.go.id/article/view/1060/jika-tidak-dikendalikan-26-juta-orang-di-dunia-menderita-kanker-.html>. Diakses tanggal 18 Maret 2015
- Johnson, Richard A and Wichern, Dean W. 1992. *Applied Multivariate Statistical Analysis*. America: Pearson International Edition.
- Lee, Y-Y., Kim, T-J., Kim, J-Y., Choi, C, H., Do, I-G., Song, S, Y., et al (2013). Genetic Profiling to Predict Recurrence of Early Cervical Cancer. *Journal Gynecologic Oncology*. 13. 650 – 654
- Lewis, M.D dan Roger, J. (2000). *An Introduction to Classification and Regression Trees (CART) Analysis*. Annual Meeting of Society For Academic Emergency. California, UCLA Medical Center
- Loka, A., Tsukuma, H., Ajiki, W., Oshima, A. (2005). Influence Of Age On Cervical Cancer Survival in Japan. *Journal Oxford JJCO*. 35. 464 – 469
- Mauliddah, N. F. (2014). *Kualitas Hidup Berdasarkan Karakteristik Pasien Kanker Serviks yang telah Melakukan Terapi di RSUD Dr. Soetomo Surabaya*. Skripsi Fakultas Kesehatan Masyarakat Universitas Jember
- Matsuura, Y., Kawagoe, T., Toki, N., Tanaka, M., Kashimura, M. (2006). Long-standing Complications after Treatment for Cancer of the Uterine Cervix – Clinical Significance of Medical Examination at 5 years after Treatment. *International Journal of Gynecological Cancer*. 16. 294 – 297

- Meilinda, T. (2014). *Kanker serviks*. Diperoleh dari: <http://mahasiswa.ung.ac.id/811412083/home/2014/12/9/malah-kanker-serviks.html>. Diakses tanggal: 27 Maret 2015
- Moeloek, FA., Nuranna, L., Wibowo, N., Purbadi, S. (2006). *Standar Pelayanan Medik Obstetri dan Ginekologi*. Jakarta: Perkumpulan Obstetri dan Ginekologi Indonesia
- Novitasari, D.A. (2014). *Analisis Survival pada Data Rekurensi dengan Menggunakan Counting Process Approach dan Model PWP-GT Study Kasus: Data Kanker Serviks di Rumah Sakit DR. Soetomo Surabaya*. Tesis Program Magister Jurusan Statistika Institut Teknologi Sepuluh Nopember
- Nurlaila, I dan Hadi, Miftachul. (2014). *Kanker: Pertumbuhan Terapi dan Nonmedis*. Diperoleh dari: <http://www.nano.lipi.go.id/utama.cgi?artikel&1187593839>. Diakses tanggal 27 Maret 2015
- Nuriyah. (2013). *Perbandingan Metode Chi-Square Automatic Interaction Detection (CHAID) dan Classification And Regression Trees (CART) dalam Menentukan Klasifikasi Alumni UIN Sunan Kalijaga Berdasarkan Masa Studi*. Skripsi Program Sarjana Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga
- Pelsi, S. (2006). *Buku Acuan Nasional Onkologi Ginekologi*. In: Aziz M Farid, Adrijojo, Saifuddin Abdul Bari, editors. Biopsi. Jakarta : Yayasan Bina Pustaka Sarwono Prawirohardjo, p. 239 – 252
- Ramdani, Y. (2013). *Klasifikasi Hasil Pap Smear Test Kanker Serviks Berdasarkan Faktor Resiko (Study Kasus di Rumah Sakit Swasta Surabaya)*. Tugas Akhir Program Sarjana Jurusan Statistika Institut Teknologi Sepuluh Nopember
- Schneider, J., 1997. *Carnegie Mellon University : School of Computer Science*. [Online] Available at: <http://www.cs.cmu.edu/~schneide/tut5/node42.html>

- Tilong, A.D. (2012). *Bebas dari Ancaman Kanker Serviks, Mengatasi dan Mencegah Penyakit Ganas dan Mematikan Bagi Kaum Wanita*. Jakarta : FlashBook
- World Health Organization. (2008). *World Cancer Report*. WHO Press
- World Health Organization. (2010). *World Cancer Report*. WHO Press
- Yamato, K., Fen, J., Kobuchi, H., Nasu, Y., Yamada, T., Nishihara, T., et al. (2006). Induction of Cell Death in Human Papillomavirus 18-Positive Cervical Cancer Cells by E6 SIRNA. *Cancer Gene Therapy*. 13: 234-241.

LAMPIRAN

Lampiran A Data Kanker Serviks

No.	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	1	47	2	3	1	2	2
2	2	44	1	3	2	1	1
3	2	46	1	3	2	1	1
4	2	54	1	3	2	2	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
271	2	56	1	3	2	1	1
272	1	48	2	2	1	1	1
273	2	52	1	3	2	1	1
274	2	50	1	2	2	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
561	1	52	1	3	4	2	1
562	2	46	1	3	2	1	1
563	2	44	1	3	2	1	1
564	2	55	1	2	2	1	1
565	2	53	1	2	2	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
806	2	53	1	2	2	1	1
807	2	60	1	2	2	1	1
808	2	46	1	3	2	1	1
809	2	51	1	3	2	1	1
810	2	27	2	3	1	1	1

Keterangan:







- Y : Status rekurensi kanker serviks
- X₁ : Usia Pasien
- X₂ : Status Anemia
- X₃ : Stadium Kanker Serviks
- X₄ : Jenis Terapi yang Dijalani Pasien
- X₅ : Status Komplikasi
- X₆ : Status Penyakit Penyerta

Lampiran B

Output CART Hasil Klasifikasi

Output CART untuk Pohon Klasifikasi Maksimal

Variable Importance

Variable	Score	
X4	100.00	
X2	71.51	
X5	31.27	
X3	23.80	
X1	13.44	
X6	11.99	

Misclassification for Learn Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	468	120	25.64	0.26
1	342	115	33.63	0.34

Misclassification for Test Data

Class	N Cases	N Mis-classed	Pct Error	Cost
2	468	130	27.78	0.28
1	342	120	35.09	0.35

Output statistik variabel penelitian

LEARNING SAMPLE VARIABLE STATISTICS

=====

VARIABLE		CLASS		OVERALL
		1	2	
Y	MEAN	1.000	2.000	1.578
	SD	0.000	0.000	0.494
	N	342.000	468.000	810.000
	SUM	342.000	936.000	1278.000
X1	MEAN	50.620	49.329	49.874
	SD	9.274	7.987	8.572
	N	342.000	468.000	810.000
	SUM	17312.000	23086.000	40398.000
X2	MEAN	1.477	1.199	1.316
	SD	0.500	0.399	0.465
	N	342.000	468.000	810.000
	SUM	505.000	561.000	1066.000
X3	MEAN	2.579	2.656	2.623
	SD	0.834	0.562	0.691
	N	342.000	468.000	810.000
	SUM	882.000	1243.000	2125.000
X4	MEAN	1.901	1.923	1.914
	SD	0.988	0.546	0.764
	N	342.000	468.000	810.000
	SUM	650.000	900.000	1550.000
X5	MEAN	1.322	1.147	1.221
	SD	0.468	0.355	0.415
	N	342.000	468.000	810.000
	SUM	452.000	537.000	989.000
X6	MEAN	1.079	1.036	1.054
	SD	0.270	0.187	0.227
	N	342.000	468.000	810.000
	SUM	369.000	485.000	854.000

Output informasi simpul (*node information*) pohon klasifikasi optimal

```
=====
NODE INFORMATION
=====
```

```
*****
*                Node 1: X4                *
*                N: 810                    *
*****
```

```
=====
= Terminal Node 1 = * Node 2 *
= N: 307 = * N: 503 *
=====
*****
```

Node 1 was split on X4
 A case goes left if X4 = (1,3,4)
 Improvement = 0.068 Complexity Threshold = 0.181

Node	Cases	Wgt Counts	Cost	Class
1	810	810.00	0.500	1
-1	307	307.00	0.278	1
2	503	503.00	0.348	2

Class	Weighted Counts		
	Top	Left	Right
1	342.00	201.00	141.00
2	468.00	106.00	362.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.500	0.722	0.348
2	0.500	0.278	0.652

Surrogate	Split	Assoc.	Improve.
1 X2	s 2	0.708	0.043
2 X5	s 2	0.365	0.021
3 X3	s 0,1,4	0.103	0.016
4 X6	s 2	0.047	0.004

Competitor	Split	Improve.
1 X2	2	0.043
2 X5	2	0.021
3 X3	0,1,4	0.016
4 X1	66.500	0.008

5 X6	2	0.004
------	---	-------

```
*****
*                               *
*           Node 2: X1         *
*           N: 503            *
*                               *
*****
```

```
*****
*      Node 3      * =      Terminal Node 5      =
*      N: 496      * =              N: 7          =
*****
```

```
Node 2 was split on X1
A case goes left if X1 <= 72.500
Improvement = 0.009      Complexity Threshold = 0.010
```

Cost	Class			
2	503	503.00	0.348	2
3	496	496.00	0.336	2
-5	7	7.00	0.000	1

Weighted Counts

Class	Top	Left	Right
1	141.00	134.00	7.00
2	362.00	362.00	0.00

Within Node Probabilities

Class	Top	Left	Right
1	0.348	0.336	1.000
2	0.652	0.664	0.000

Competitor	Split	Improve.
1 X5	2	0.005
2 X2	2	0.005
3 X6	2	0.003
4 X3	0,4	0.002

```
*****
*                               *
*           Node 3: X2          *
*           N: 496              *
*                               *
*****
```

===== *


```

= Terminal Node 2 = * Node 4 *
= N: 19 = * N: 477 *
===== *****

```

Node 3 was split on X2

A case goes left if X2 = (2)

Improvement = 0.005 Complexity Threshold = 0.008

Node	Cases	Wgt	Counts	Cost	Class
3	496		496.00	0.336	2
-2	19		19.00	0.347	1
4	477		477.00	0.322	2

Class	Weighted Counts		
	Top	Left	Right
1	134.00	11.00	123.00
2	362.00	8.00	354.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.336	0.653	0.322
2	0.664	0.347	0.678

Surrogate	Split	Assoc.	Improve.
1 X3	s 0	0.016	.296561E-03

Competitor	Split	Improve.
1 X5	2	0.004
2 X6	2	0.003
3 X3	0,4	0.002
4 X1	54.500	0.002

```

*****
* Node 4: X6 *
* N: 477 *
*****

```

```

=====
= Terminal Node 3 = = Terminal Node 4 =
= N: 14 = = N: 463 =
=====

```

Node 4 was split on X6

A case goes left if X6 = (2)

Improvement = 0.004 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
4	477		477.00	0.322	2
-3	14		14.00	0.354	1
-4	463		463.00	0.311	2

Class	Weighted Counts		
	Top	Left	Right
1	123.00	8.00	115.00
2	354.00	6.00	348.00

Class	Within Node Probabilities		
	Top	Left	Right
1	0.322	0.646	0.311
2	0.678	0.354	0.689

Surrogate	Split	Assoc.	Improve.
1 X1 s	27.500	0.059	.222296E-03

Competitor	Split	Improve.
1 X1	61.500	0.002
2 X3	4	0.002
3 X5	2	0.001

Lampiran C

Output *Terminal Nodes* Untuk Klasifikasi Data *Learning*

Output CART untuk *Fold 1*

```
=====
  TERMINAL NODE INFORMATION
=====
```

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity
								Threshold
1	17	0.031	1	0.000	1	17	1.000	0.001
					2	0	0.000	
2	2	0.003	2	0.000	1	0	0.000	0.002
					2	2	1.000	
3	41	0.070	1	0.190	1	31	0.810	0.002
					2	10	0.190	
4	2	0.004	1	0.000	1	2	1.000	0.001
					2	0	0.000	
5	4	0.007	1	0.195	1	3	0.805	0.001
					2	1	0.195	
6	14	0.023	1	0.288	1	9	0.712	0.001
					2	5	0.288	
7	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
8	7	0.010	2	0.355	1	2	0.355	0.002
					2	5	0.645	
9	6	0.009	2	0.407	1	2	0.407	0.002
					2	4	0.593	
10	6	0.008	2	0.216	1	1	0.216	0.003
					2	5	0.784	
11	6	0.011	1	0.000	1	6	1.000	0.001
					2	0	0.000	
12	25	0.041	1	0.32	1	15	0.673	0.001
					2	10	0.327	
13	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
14	24	0.042	1	0.127	1	20	0.873	.6666E-03
					2	4	0.127	
15	8	0.014	1	0.195	1	6	0.805	0.002
					2	2	0.195	
16	11	0.018	1	0.378	1	6	0.622	0.003
					2	5	0.378	
17	4	0.006	2	0.314	1	1	0.314	.1611E-03

					2	3	0.686	
18	7	0.011	1	0.493	1	3	0.507	.1611E-03
					2	4	0.493	
19	1	0.001	2	0.000	1	0	0.000	.8058E-04
					2	1	1.000	
20	3	0.005	1	0.000	1	3	1.000	0.002
					2	0	0.000	
21	7	0.012	1	0.226	1	5	0.774	0.004
					2	2	0.226	
22	5	0.007	2	0.256	1	1	0.256	0.004
					2	4	0.744	
23	2	0.003	2	0.000	1	0	0.000	0.002
					2	2	1.000	
24	8	0.015	1	0.000	1	8	1.000	0.001
					2	0	0.000	
25	9	0.016	1	0.083	1	8	0.917	0.001
					2	1	0.083	
26	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
27	9	0.013	2	0.282	1	2	0.282	0.006
					2	7	0.718	
28	19	0.031	1	0.298	1	12	0.702	0.001
					2	7	0.298	
29	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
30	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
31	3	0.005	1	0.000	1	3	1.000	0.003
					2	0	0.000	
32	9	0.013	2	0.282	1	2	0.282	0.004
					2	7	0.718	
33	2	0.004	1	0.000	1	2	1.000	0.004
					2	0	0.000	
34	2	0.004	1	0.000	1	2	1.000	0.004
					2	0	0.000	
35	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
36	7	0.010	2	0.355	1	2	0.355	.2494E-03
					2	5	0.645	
37	2	0.003	1	0.421	1	1	0.579	.4981E-03
					2	1	0.421	
38	10	0.013	2	0.000	1	0	0.000	.4981E-03
					2	10	1.000	
39	11	0.015	2	0.000	1	0	0.000	.2494E-03

					2	11	1.000	
40	2	0.003	1	0.421	1	1	0.579	.4981E-03
					2	1	0.421	
41	9	0.012	2	0.147	1	1	0.147	.4981E-03
					2	8	0.853	
42	6	0.009	1	0.421	1	3	0.579	0.001
					2	3	0.421	
43	8	0.013	1	0.421	1	4	0.579	.8351E-03
					2	4	0.421	
44	3	0.004	2	0.407	1	1	0.407	.8351E-03
					2	2	0.593	
45	9	0.012	2	0.147	1	1	0.147	.9963E-03
					2	8	0.853	
46	5	0.008	1	0.327	1	3	0.673	0.001
					2	2	0.327	
47	8	0.012	2	0.452	1	3	0.452	0.001
					2	5	0.548	
48	29	0.041	2	0.222	1	5	0.222	.9641E-03
					2	24	0.778	
49	14	0.019	2	0.096	1	1	0.096	.80341E-03
					2	13	0.904	
50	12	0.019	1	0.342	1	7	0.658	0.006
					2	5	0.342	
51	25	0.037	2	0.393	1	8	0.393	0.006
					2	17	0.607	
52	17	0.024	2	0.155	1	2	0.155	.8560E-03
					2	15	0.845	
53	24	0.036	2	0.407	1	8	0.407	.9987E-03
					2	16	0.593	
54	8	0.012	2	0.314	1	2	0.314	0.004
					2	6	0.686	
55	4	0.007	1	0.195	1	3	0.805	0.004
					2	1	0.195	
56	25	0.036	2	0.303	1	6	0.303	0.002
					2	19	0.697	
57	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
58	3	0.004	2	0.000	1	0	0.000	0.001
					2	3	1.000	
59	38	0.053	2	0.172	1	5	0.172	.3985E-03
					2	33	0.828	
60	12	0.017	2	0.314	1	3	0.314	.6642E-03
					2	9	0.686	
61	8	0.013	1	0.421	1	4	0.579	0.001

					2	4	0.421	
62	8	0.012	2	0.452	1	3	0.452	0.001
					2	5	0.548	
63	1	0.001	2	0.000	1	0	0.000	.9963E-03
					2	1	1.000	
64	24	0.034	2	0.216	1	4	0.216	.4981E-03
					2	20	0.784	
65	6	0.009	1	0.421	1	3	0.579	.8351E-03
					2	3	0.421	
66	6	0.009	2	0.407	1	2	0.407	.6593E-03
					2	4	0.593	
67	9	0.014	1	0.476	1	4	0.524	.6593E-03
					2	5	0.476	
68	1	0.001	2	0.000	1	0	0.000	.7179E-03
					2	1	1.000	
69	5	0.007	2	0.256	1	1	0.256	.5384E-03
					2	4	0.744	
70	2	0.003	2	0.000	1	0	0.000	.4307E-03
					2	2	1.000	
71	8	0.011	2	0.000	1	0	0.000	0.001
					2	8	1.000	
72	9	0.016	1	0.083	1	8	0.917	0.015
					2	1	0.083	

Output CART untuk Fold 2

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	12	0.022	1	0.000	1	12	1.000	6666E-03
					2	0	0.000	
2	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
3	14	0.025	1	0.108	1	12	0.892	0.001
					2	2	0.108	
4	9	0.016	1	0.000	1	9	1.000	7228E-03
					2	0	0.000	
5	4	0.007	1	0.000	1	4	1.000	0.001
					2	0	0.000	
6	4	0.006	2	0.314	1	1	0.314	0.002

					2	3	0.686	
7	4	0.007	1	0.000	1	4	1.000	.6667E-03
					2	0	0.000	
8	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
9	10	0.017	1	0.154	1	8	0.846	0.001
					2	2	0.154	
10	8	0.012	2	0.452	1	3	0.452	0.001
					2	5	0.548	
11	30	0.052	1	0.127	1	25	0.873	0.001
					2	5	0.127	
12	5	0.008	2	0.478	1	2	0.478	.7545E-03
					2	3	0.522	
13	6	0.009	2	0.407	1	2	0.407	0.002
					2	4	0.593	
14	10	0.017	1	0.154	1	8	0.846	0.002
					2	2	0.154	
15	3	0.005	1	0.000	1	3	1.000	0.003
					2	0	0.000	
16	5	0.007	2	0.256	1	1	0.256	0.004
					2	4	0.744	
17	8	0.013	1	0.304	1	5	0.696	0.004
					2	3	0.304	
18	3	0.004	2	0.000	1	0	0.000	0.003
					2	3	1.000	
19	9	0.016	1	0.000	1	9	1.000	0.001
					2	0	0.000	
20	5	0.009	1	0.154	1	4	0.846	0.002
					2	1	0.154	
21	8	0.013	1	0.304	1	5	0.696	0.003
					2	3	0.304	
22	2	0.003	2	0.000	1	0	0.000	0.003
					2	2	1.000	
23	2	0.003	2	0.000	1	0	0.000	0.003
					2	2	1.000	
24	6	0.011	1	0.000	1	6	1.000	0.001
					2	0	0.000	
25	20	0.033	1	0.282	1	13	0.718	0.004
					2	7	0.282	
26	5	0.007	2	0.256	1	1	0.256	0.004
					2	4	0.744	
27	6	0.010	1	0.127	1	5	0.873	0.002
					2	1	0.127	
28	11	0.016	2	0.340	1	3	0.340	0.003

					2	8	0.660	
29	4	0.006	2	0.314	1	1	0.314	0.002
					2	3	0.686	
30	9	0.016	1	0.083	1	8	0.917	0.002
					2	1	0.083	
31	7	0.012	1	0.226	1	5	0.774	0.003
					2	2	0.226	
32	2	0.004	1	0.000	1	2	1.000	0.004
					2	0	0.000	
33	11	0.016	2	0.340	1	3	0.340	0.004
					2	8	0.660	
34	6	0.008	2	0.216	1	1	0.216	0.002
					2	5	0.784	
35	9	0.014	1	0.36	1	5	0.632	0.004
					2	4	0.368	
36	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
37	10	0.017	1	0.238	1	7	0.762	0.001
					2	3	0.238	
38	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
39	16	0.021	2	0.000	1	0	0.000	0.002
					2	16	1.000	
40	2	0.004	1	0.000	1	2	1.000	0.004
					2	0	0.000	
41	2	0.003	2	0.000	1	0	0.000	.7472E-03
					2	2	1.000	
42	10	0.015	2	0.371	1	3	0.371	.9963E-03
					2	7	0.629	
43	4	0.006	1	0.421	1	2	0.579	.9963E-03
					2	2	0.421	
44	6	0.008	2	0.216	1	1	0.216	.4982E-03
					2	5	0.784	
45	6	0.009	1	0.421	1	3	0.579	0.001
					2	3	0.421	
46	43	0.059	2	0.093	1	3	0.093	0.001
					2	40	0.907	
47	4	0.007	1	0.195	1	3	0.805	0.004
					2	1	0.195	
48	2	0.003	1	0.421	1	1	0.579	.4981E-03
					2	1	0.421	
49	13	0.019	2	0.379	1	4	0.379	.4981E-03
					2	9	0.621	
50	4	0.006	1	0.421	1	2	0.579	0.000

				2	2	0.421	
51	7	0.011	1	0.493	1	3	0.507 0.000
				2	4	0.493	
52	1	0.001	2	0.000	1	0	0.000 0.001
				2	1	1.000	
53	51	0.074	2	0.320	1	13	0.320 .4139E-03
				2	38	0.680	
54	63	0.091	2	0.282	1	14	0.282 .3311E-03
				2	49	0.718	
55	31	0.043	2	0.169	1	4	0.169 .2759E-03
				2	27	0.831	
56	8	0.013	1	0.421	1	4	0.579 0.001
				2	4	0.421	
57	13	0.019	2	0.292	1	3	0.292 .6593E-03
				2	10	0.708	
58	9	0.014	1	0.476	1	4	0.524 .6593E-03
				2	5	0.476	
59	13	0.019	2	0.379	1	4	0.379 .8840E-03
				2	9	0.621	
60	1	0.001	2	0.000	1	0	0.000 .6630E-03
				2	1	1.000	
61	1	0.001	2	0.000	1	0	0.000 .5304E-03
				2	1	1.000	
62	1	0.001	2	0.000	1	0	0.000 .4420E-03
				2	1	1.000	
63	27	0.038	2	0.193	1	4	0.193 .3315E-03
				2	23	0.807	
64	3	0.005	1	0.267	1	2	0.733 0.002
				2	1	0.267	
65	10	0.015	2	0.371	1	3	0.371 .1611E-03
				2	7	0.629	
66	7	0.011	1	0.493	1	3	0.507 .1611E-03
				2	4	0.493	
67	4	0.006	2	0.314	1	1	0.314 .8058E-04
				2	3	0.686	
68	1	0.001	2	0.000	1	0	0.000 .5372E-04
				2	1	1.000	
69	1	0.001	2	0.000	1	0	0.000 .4029E-04
				2	1	1.000	
70	1	0.001	2	0.000	1	0	0.000 .3223E-04
				2	1	1.000	
71	4	0.006	1	0.421	1	2	0.579 .9963E-03
				2	2	0.421	
72	1	0.001	2	0.000	1	0	0.000 .4981E-03

```

2      1  1.000
73  7   0.013   1   0.000   1   7  1.000   0.013
2      0  0.000

```

Output CART untuk Fold 3

```

=====
  TERMINAL NODE INFORMATION
=====

```

Node	N	Prob	Class	Cost	Class	N	Complexity Prob Threshold
1	10	0.018	1	0.000	1	10	1.000 .6684E-03
					2	0	0.000
2	1	0.001	2	0.000	1	0	0.000 0.001
					2	1	1.000
3	14	0.025	1	0.109	1	12	0.891 0.001
					2	2	0.109
4	8	0.014	1	0.196	1	6	0.804 0.002
					2	2	0.196
5	6	0.009	2	0.406	1	2	0.406 0.002
					2	4	0.594
6	39	0.068	1	0.118	1	33	0.882 0.002
					2	6	0.118
7	4	0.006	2	0.313	1	1	0.313 0.002
					2	3	0.687
8	5	0.008	1	0.328	1	3	0.672 0.003
					2	2	0.328
9	10	0.014	2	0.254	1	2	0.254 0.003
					2	8	0.746
10	4	0.007	1	0.000	1	4	1.000 .4456E-03
					2	0	0.000
11	2	0.004	1	0.000	1	2	1.000 .1805E-03
					2	0	0.000
12	9	0.014	1	0.370	1	5	0.630 .3610E-03
					2	4	0.370
13	5	0.008	2	0.476	1	2	0.476 .3610E-03
					2	3	0.524
14	9	0.015	1	0.173	1	7	0.827 0.001
					2	2	0.173
15	1	0.001	2	0.000	1	0	0.000 0.001
					2	1	1.000

16	8	0.013	1	0.423	1	4	0.577	.3342E-03
					2	4	0.423	
17	3	0.005	1	0.268	1	2	0.732	0.001
					2	1	0.268	
18	8	0.013	1	0.423	1	4	0.577	0.002
					2	4	0.423	
19	2	0.003	2	0.000	1	0	0.000	0.002
					2	2	1.000	
20	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
21	27	0.046	1	0.173	1	21	0.827	.8021E-03
					2	6	0.173	
22	7	0.012	1	0.109	1	6	0.891	0.002
					2	1	0.109	
23	7	0.012	1	0.227	1	5	0.773	0.003
					2	2	0.227	
24	8	0.012	2	0.450	1	3	0.450	0.001
					2	5	0.550	
25	10	0.017	1	0.155	1	8	0.845	0.001
					2	2	0.155	
26	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
27	12	0.019	1	0.344	1	7	0.656	0.001
					2	5	0.344	
28	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
29	12	0.017	2	0.214	1	2	0.214	0.005
					2	10	0.786	
30	2	0.003	2	0.000	1	0	0.000	0.003
					2	2	1.000	
31	13	0.023	1	0.118	1	11	0.882	0.002
					2	2	0.118	
32	9	0.013	2	0.406	1	3	0.406	0.003
					2	6	0.594	
33	8	0.014	1	0.095	1	7	0.905	0.003
					2	1	0.095	
34	12	0.019	1	0.344	1	7	0.656	0.006
					2	5	0.344	
35	2	0.004	1	0.000	1	2	1.000	0.004
					2	0	0.000	
36	8	0.013	1	0.423	1	4	0.577	0.002
					2	4	0.423	
37	9	0.012	2	0.000	1	0	0.000	0.002
					2	9	1.000	

38	4	0.005	2	0.000	1	0	0.000	.9124E-03
					2	4	1.000	
39	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
40	12	0.017	2	0.110	1	1	0.110	0.002
					2	11	0.890	
41	15	0.020	2	0.000	1	0	0.000	.6082E-03
					2	15	1.000	
42	9	0.014	1	0.478	1	4	0.522	.8698E-03
					2	5	0.478	
43	19	0.026	2	0.138	1	2	0.138	.9124E-03
					2	17	0.862	
44	6	0.010	1	0.268	1	4	0.732	0.005
					2	2	0.268	
45	22	0.032	2	0.286	1	5	0.286	0.005
					2	17	0.714	
46	10	0.013	2	0.000	1	0	0.000	.2439E-03
					2	10	1.000	
47	2	0.003	1	0.423	1	1	0.577	.4879E-03
					2	1	0.423	
48	15	0.022	2	0.332	1	4	0.332	.4879E-03
					2	11	0.668	
49	5	0.008	1	0.328	1	3	0.672	0.003
					2	2	0.328	
50	7	0.010	2	0.353	1	2	0.353	0.003
					2	5	0.647	
51	26	0.039	2	0.378	1	8	0.378	0.001
					2	18	0.622	
52	1	0.001	2	0.000	1	0	0.000	.9335E-03
					2	1	1.000	
53	20	0.028	2	0.132	1	2	0.132	.5601E-03
					2	18	0.868	
54	9	0.013	2	0.406	1	3	0.406	0.003
					2	6	0.594	
55	5	0.008	1	0.328	1	3	0.672	0.003
					2	2	0.328	
56	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
57	7	0.010	2	0.185	1	1	0.185	.9335E-03
					2	6	0.815	
58	49	0.070	2	0.259	1	10	0.259	.7001E-03
					2	39	0.741	

59	8	0.013	1	0.423	1	4	0.577	0.002
					2	4	0.423	
60	19	0.028	2	0.386	1	6	0.386	0.002
					2	13	0.614	
61	12	0.018	2	0.313	1	3	0.313	.9758E-03
					2	9	0.687	
62	1	0.001	2	0.000	1	0	0.000	.6505E-03
					2	1	1.000	
63	3	0.004	2	0.000	1	0	0.000	.8251E-03
					2	3	1.000	
64	21	0.030	2	0.185	1	3	0.185	.7426E-03
					2	18	0.815	
65	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
66	2	0.003	1	0.423	1	1	0.577	.4879E-03
					2	1	0.423	
67	10	0.015	2	0.476	1	4	0.476	.4879E-03
					2	6	0.524	
68	12	0.018	2	0.406	1	4	0.406	.2439E-03
					2	8	0.594	
69	1	0.001	2	0.000	1	0	0.000	.1626E-03
					2	1	1.000	
70	2	0.004	1	0.000	1	2	1.000	0.002
					2	0	0.000	
71	3	0.004	2	0.000	1	0	0.000	0.002
					2	3	1.000	
72	4	0.005	2	0.000	1	0	0.000	0.001
					2	4	1.000	
73	7	0.013	1	0.000	1	7	1.000	0.013
					2	0	0.000	

Output CART untuk Fold 4

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity Threshold
1	12	0.022	1	0.000	1	12	1.000	0.003
					2	0	0.000	
2	3	0.004	2	0.406	1	1	0.406	.8489E-03
					2	2	0.594	
3	22	0.040	1	0.034	1	21	0.966	.6684E-03
					2	1	0.034	

4	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
5	15	0.026	1	0.101	1	13	0.899	0.001
					2	2	0.101	
6	13	0.022	1	0.246	1	9	0.754	.4456E-03
					2	4	0.246	
7	3	0.004	2	0.406	1	1	0.406	.8489E-03
					2	2	0.594	
8	4	0.006	2	0.313	1	1	0.313	0.002
					2	3	0.687	
9	3	0.005	1	0.000	1	3	1.000	0.005
					2	0	0.000	
10	8	0.011	2	0.000	1	0	0.000	0.005
					2	8	1.000	
11	16	0.028	1	0.095	1	14	0.905	0.001
					2	2	0.095	
12	9	0.015	1	0.173	1	7	0.827	0.001
					2	2	0.173	
13	11	0.018	1	0.295	1	7	0.705	0.001
					2	4	0.295	
14	5	0.008	1	0.328	1	3	0.672	0.002
					2	2	0.328	
15	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
16	3	0.005	1	0.268	1	2	0.732	0.002
					2	1	0.268	
17	5	0.008	1	0.328	1	3	0.672	0.003
					2	2	0.328	
18	9	0.013	2	0.281	1	2	0.281	0.003
					2	7	0.719	
19	8	0.013	1	0.423	1	4	0.577	0.002
					2	4	0.423	
20	5	0.007	2	0.254	1	1	0.254	0.002
					2	4	0.746	
21	7	0.013	1	0.000	1	7	1.000	.4244E-03
					2	0	0.000	
22	16	0.028	1	0.145	1	13	0.855	.8489E-03
					2	3	0.145	
23	3	0.004	2	0.406	1	1	0.406	.8489E-03
					2	2	0.594	
24	23	0.038	1	0.320	1	14	0.680	0.007
					2	9	0.320	
25	5	0.007	2	0.000	1	0	0.000	0.007
					2	5	1.000	

26	4	0.007	1	0.000	1	4	1.000	.8912E-03
					2	0	0.000	
27	12	0.021	1	0.128	1	10	0.872	0.001
					2	2	0.128	
28	8	0.013	1	0.305	1	5	0.695	0.003
					2	3	0.305	
29	2	0.003	2	0.000	1	0	0.000	0.003
					2	2	1.000	
30	4	0.007	1	0.000	1	4	1.000	0.001
					2	0	0.000	
31	8	0.012	2	0.450	1	3	0.450	0.001
					2	5	0.550	
32	2	0.004	1	0.000	1	2	1.000	0.001
					2	0	0.000	
33	8	0.012	2	0.450	1	3	0.450	0.001
					2	5	0.550	
34	2	0.004	1	0.000	1	2	1.000	0.004
					2	0	0.000	
35	3	0.005	1	0.268	1	2	0.732	0.002
					2	1	0.268	
36	9	0.012	2	0.000	1	0	0.000	0.002
					2	9	1.000	
37	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
38	2	0.003	1	0.423	1	1	0.577	.4879E-03
					2	1	0.423	
39	11	0.016	2	0.233	1	2	0.233	.4879E-03
					2	9	0.767	
40	6	0.010	1	0.268	1	4	0.732	.3610E-03
					2	2	0.268	
41	5	0.008	2	0.476	1	2	0.476	.3610E-03
					2	3	0.524	
42	7	0.010	2	0.185	1	1	0.185	0.002
					2	6	0.815	
43	7	0.011	1	0.355	1	4	0.645	0.002
					2	3	0.355	
44	33	0.047	2	0.233	1	6	0.233	.1268E-03
					2	27	0.767	
45	14	0.022	1	0.494	1	6	0.506	.2537E-03
					2	8	0.494	
46	7	0.009	2	0.000	1	0	0.000	0.000
					2	7	1.000	
47	19	0.028	2	0.328	1	5	0.328	0.000
					2	14	0.672	

48	11	0.016	2	0.339	1	3	0.339	0.000
					2	8	0.661	
49	8	0.013	1	0.423	1	4	0.577	0.002
					2	4	0.423	
50	41	0.055	2	0.033	1	1	0.033	0.002
					2	40	0.967	
51	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
52	17	0.025	2	0.363	1	5	0.363	0.001
					2	12	0.637	
53	5	0.008	1	0.328	1	3	0.672	0.002
					2	2	0.328	
54	6	0.009	2	0.406	1	2	0.406	0.002
					2	4	0.594	
55	21	0.030	2	0.185	1	3	0.185	0.001
					2	18	0.815	
56	6	0.010	1	0.268	1	4	0.732	0.005
					2	2	0.268	
57	14	0.020	2	0.271	1	3	0.271	0.005
					2	11	0.729	
58	8	0.011	2	0.163	1	1	0.163	.4879E-03
					2	7	0.837	
59	4	0.006	1	0.423	1	2	0.577	.9758E-03
					2	2	0.423	
60	18	0.026	2	0.281	1	4	0.281	.9758E-03
					2	14	0.719	
61	65	0.093	2	0.236	1	12	0.236	.3252E-03
					2	53	0.764	
62	3	0.004	2	0.000	1	0	0.000	.2439E-03
					2	3	1.000	
63	15	0.024	1	0.328	1	9	0.672	0.001
					2	6	0.328	
64	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
65	6	0.009	2	0.406	1	2	0.406	0.002
					2	4	0.594	
66	2	0.004	1	0.000	1	2	1.000	0.001
					2	0	0.000	
67	4	0.006	2	0.313	1	1	0.313	0.002
					2	3	0.687	
68	3	0.004	2	0.000	1	0	0.000	0.003
					2	3	1.000	
69	5	0.009	1	0.000	1	5	1.000	0.002
					2	0	0.000	

Output CART untuk Fold 5

=====

TERMINAL NODE INFORMATION

=====

Node	N	Prob	Class	Cost	Class	N	Prob	Complexity
								Threshold
1	21	0.036	1	0.186	1	16	0.814	0.001
					2	5	0.186	
2	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
3	19	0.034	1	0.039	1	18	0.961	.6684E-03
					2	1	0.039	
4	4	0.006	2	0.313	1	1	0.313	0.002
					2	3	0.687	
5	16	0.029	1	0.047	1	15	0.953	.6684E-03
					2	1	0.047	
6	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
7	33	0.057	1	0.140	1	27	0.860	0.001
					2	6	0.140	
8	2	0.003	2	0.000	1	0	0.000	0.003
					2	2	1.000	
9	5	0.009	1	0.155	1	4	0.845	0.004
					2	1	0.155	
10	5	0.007	2	0.254	1	1	0.254	0.004
					2	4	0.746	
11	14	0.023	1	0.355	1	8	0.645	0.004
					2	6	0.355	
12	16	0.023	2	0.240	1	3	0.240	0.003
					2	13	0.760	
13	15	0.026	1	0.101	1	13	0.899	0.002
					2	2	0.101	
14	7	0.012	1	0.227	1	5	0.773	0.001
					2	2	0.227	
15	8	0.013	1	0.305	1	5	0.695	0.001
					2	3	0.305	
16	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
17	5	0.008	1	0.328	1	3	0.672	0.003
					2	2	0.328	
18	5	0.007	2	0.254	1	1	0.254	0.003
					2	4	0.746	

19	2	0.003	2	0.000	1	0	0.000	0.002
					2	2	1.000	
20	11	0.020	1	0.000	1	11	1.000	0.001
					2	0	0.000	
21	9	0.015	1	0.173	1	7	0.827	0.003
					2	2	0.173	
22	2	0.003	2	0.000	1	0	0.000	0.003
					2	2	1.000	
23	8	0.013	1	0.305	1	5	0.695	0.002
					2	3	0.305	
24	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
25	5	0.008	1	0.328	1	3	0.672	0.003
					2	2	0.328	
26	7	0.010	2	0.185	1	1	0.185	0.003
					2	6	0.815	
27	9	0.015	1	0.268	1	6	0.732	0.003
					2	3	0.268	
28	7	0.010	2	0.353	1	2	0.353	0.003
					2	5	0.647	
29	5	0.007	2	0.254	1	1	0.254	0.002
					2	4	0.746	
30	2	0.003	2	0.000	1	0	0.000	0.002
					2	2	1.000	
31	8	0.014	1	0.095	1	7	0.905	0.002
					2	1	0.095	
32	5	0.009	1	0.000	1	5	1.000	0.002
					2	0	0.000	
33	9	0.015	1	0.268	1	6	0.732	0.001
					2	3	0.268	
34	1	0.001	2	0.000	1	0	0.000	0.001
					2	1	1.000	
35	3	0.005	1	0.268	1	2	0.732	0.002
					2	1	0.268	
36	7	0.010	2	0.185	1	1	0.185	0.002
					2	6	0.815	
37	11	0.015	2	0.000	1	0	0.000	.4879E-03
					2	11	1.000	
38	2	0.003	1	0.423	1	1	0.577	.4879E-03
					2	1	0.423	
39	2	0.003	1	0.423	1	1	0.577	.4879E-03
					2	1	0.423	
40	11	0.016	2	0.233	1	2	0.233	.4879E-03
					2	9	0.767	

41	6	0.009	2	0.406	1	2	0.406	.9758E-03
					2	4	0.594	
42	4	0.006	1	0.423	1	2	0.577	.9758E-03
					2	2	0.423	
43	1	0.001	2	0.000	1	0	0.000	.4879E-03
					2	1	1.000	
44	9	0.013	2	0.146	1	1	0.146	.3252E-03
					2	8	0.854	
45	4	0.006	1	0.423	1	2	0.577	.9758E-03
					2	2	0.423	
46	8	0.013	1	0.423	1	4	0.577	0.002
					2	4	0.423	
47	35	0.047	2	0.000	1	0	0.000	0.002
					2	35	1.000	
48	5	0.008	1	0.328	1	3	0.672	0.003
					2	2	0.328	
49	1	0.002	1	0.000	1	1	1.000	0.002
					2	0	0.000	
50	34	0.049	2	0.296	1	8	0.296	0.002
					2	26	0.704	
51	4	0.007	1	0.000	1	4	1.000	0.002
					2	0	0.000	
52	6	0.009	2	0.406	1	2	0.406	0.002
					2	4	0.594	
53	22	0.032	2	0.286	1	5	0.286	0.004
					2	17	0.714	
54	16	0.022	2	0.083	1	1	0.083	0.002
					2	15	0.917	
55	9	0.013	2	0.406	1	3	0.406	0.003
					2	6	0.594	
56	7	0.011	1	0.355	1	4	0.645	0.003
					2	3	0.355	
57	17	0.026	2	0.489	1	7	0.489	0.000
					2	10	0.511	
58	27	0.039	2	0.281	1	6	0.281	0.000
					2	21	0.719	
59	7	0.010	2	0.353	1	2	0.353	0.003
					2	5	0.647	
60	4	0.007	1	0.196	1	3	0.804	0.003
					2	1	0.196	
61	4	0.005	2	0.000	1	0	0.000	0.002
					2	4	1.000	
62	26	0.037	2	0.199	1	4	0.199	.12685-03
					2	22	0.801	

63	7	0.011	1	0.494	1	3	0.506	.1268E-03
					2	4	0.494	
64	25	0.035	2	0.206	1	4	0.206	.1626E-03
					2	21	0.794	
65	2	0.003	1	0.423	1	1	0.577	.4879E-03
					2	1	0.423	
66	8	0.012	2	0.313	1	2	0.313	.4879E-03
					2	6	0.687	
67	11	0.016	2	0.233	1	2	0.233	.2439E-03
					2	9	0.767	
68	2	0.003	2	0.000	1	0	0.000	.8275E-03
					2	2	1.000	
69	3	0.005	1	0.268	1	2	0.732	0.002
					2	1	0.268	
70	6	0.009	2	0.406	1	2	0.406	.1268E-03
					2	4	0.594	
71	7	0.011	1	0.494	1	3	0.506	.1268E-03
					2	4	0.494	
72	2	0.003	2	0.000	1	0	0.000	0.001
					2	2	1.000	
73	4	0.005	2	0.000	1	0	0.000	0.001
					2	4	1.000	
74	10	0.017	1	0.155	1	8	0.845	0.012
					2	2	0.155	

BIODATA PENULIS



Penulis, Siti Holis Sumartini, lahir di Pamekasan, 14 Juli 1992. Penulis merupakan anak keempat dari empat bersaudara. Penulis telah menempuh pendidikan formal yaitu di TK Muslimat NU, SDN Bunder II Pamekasan, SMPN I Pademawu dan SMAN I Pamekasan. Setelah lulus dari SMAN tahun 2010, penulis

mengikuti tes diploma III reguler dan diterima di Jurusan Statistika FMIPA-ITS tahun 2010 dan menjadi keluarga sigma 21. Setelah lulus diploma III di Jurusan Statistika ITS tahun 2013, kemudian penulis melanjutkan kuliah Lintas Jalur di Jurusan Statistika ITS. Selama menjadi Mahasiswa, penulis menjadi salah satu anggota divisi Pst tahun 2011-2012. Dengan motto hidup “*you can when you believe*” membuat penulis selalu optimis dan berpikir positif untuk selalu berusaha melakukan yang terbaik untuk setiap hal. Kesadaran bahwa setiap orang merupakan tempat untuk berbeda-beda serta belajar untuk sabar dan tidak apatis. Bagi pembaca yang memiliki saran, kritik dan lain sebagainya dapat disampaikan melalui email : holissumartini14@gmail.com